SNP DISCOVERY AND MAPPING QTLS ASSOCIATED WITH ROOT TRAITS AND MORPHOLOGICAL TRAITS IN TOMATO

A Thesis

by

LIMENG XIE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	John L. Jifon
Co-Chair of Committee,	Kevin M. Crosby
Committee Member,	Wayne Smith
Head of Department,	Daniel R. Lineberger

December 2016

Major Subject: Horticulture

ABSTRACT

Tomato (*Solanum lycopersicum* L.) is an economically important vegetable crop worldwide and like many other crops, its productivity is severely impacted by abiotic stresses. Establishment of an efficient root system is a means to minimize these negative impacts under stress environments, especially in a low nutrient and drought-prone environment. However, there is little information about genetic control of root traits in tomato. In our study, a segregating F2 population derived from a cross between an advanced breeding line RvT1 (*Solanum lycopersicum*) and a wild species *Solanum cheesmaniae* was used to map root traits and other morphological traits, which allowed the study of the genetic basis for several root traits in tomato.

We applied Genotyping-By-Sequencing (GBS) methodology to discover single nucleotide polymorphisms (SNPs) for constructing a linkage map, which was composed of 742 SNPs and covered 1319.47 cM with an average distance of 1.78 cM between adjacent markers. We have identified 27 QTLs for the root and shoot traits. One common QTL (flanking by chr04_3261417) associated with root length, root surface area, root volume, root fresh weight and root dry weight was identified, which could be a useful marker to screen these traits simultaneously. Our results suggested root traits were regulated by several major QTLs and a suite of small-effect QTLs.

DEDICATION

This thesis is dedicated to my parents, fiancé and beloved friends.

ACKNOWLEDGEMENTS

I am sincerely grateful to my two major advisors Dr. John Jifon and Dr. Kevin Crosby. Dr. Jifon gave me the opportunity to pursue my master degree at Texas A&M University and provided valuable advice for this project. Dr. Crosby was very helpful at every aspect. When I encountered difficulties, he was always willing and efficient to help me to solve the problems. Without their support, patience and devotion, I couldn't finish this project. I am also very thankful for Dr. Wayne Smith for agreeing to serve as my committee member. It was very pleasant to take his plant breeding class.

Special gratitude goes to Dr. Patricia Klein for her help with analyzing the sequencing data and being patient to explain genetic concept to me. I also appreciate Mrs. Natalie Paterson and Nolan Bentley with their assistance in the GBS experiments.

I would like to thank Stella Kang for her help in root phenotyping work. I am also grateful to my colleagues of Dr. Crosby lab, Stephen Perry, Naining Chi and Teresa McIntyre for their assistance in the greenhouse work.

Thank you to the horticulture department community for being so friendly and accommodative of me. And thank to my close friends, Tingting Zhao and Jisun Lee for their accompany and friendship during my time at College Station.

I wish to thank my parents for their forever love, encouragement and support. I am so happy and lucky to have them as my parents.

Lastly and most importantly, I would like to thank my fiancé, Huang Chen, for his patience, love and support during the time we are together. Also thank for his help to revise the format of this thesis.

NOMENCLATURE

ANOVA	Analysis of Variance
GBS	Genotyping By Sequencing
LOD	Logarithm of Odd
QTL	Quantitative Trait Loci
RIL	Recombinant Inbred Line
SNP	Single Nucleotide Polymorphism

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

GENERAL INTRODUCTION

The world tomato production statistics

The cultivated tomato (*Solanum lycopersicum*) is one of broadly produced consumed vegetables worldwide. The cultivated tomato industries are divided into fresh-market and processing food industries. They are consumed primarily as a vegetable or fruit in the fresh market. In the processing market, they are used for seasoning and flavoring, like ketchup, tomato paste, salsa, etc. Occasionally, tomatoes are also used as landscape plants (ornamental value) (Heuvelink, 2005).

In 2013, the annual worldwide production of tomatoes was estimated at 163.96 million tones, which had increased by 400% in the last five decades (http://faostat3.fao.org/browse/Q/QC/E, accessed on 8/16/2015). Meanwhile, the annual yield of tomatoes went up from 16,434.6 Kg/Ha in 1961 to 34,698.3 Kg/Ha in 2013 (FAOSTAT, 2013). These figures may still underestimate the real production weight and yield, considering that tomatoes are also grown in private fields and gardens throughout the tropics and subtropics (Liedl et. al, 2013). This part of production is consumed locally and serves as an important food source for the population (Liedl et al., 2013). Of all the global tomato production, 60.5% was distributed in the Asia; While 15% of the production was in America in 2013 (http://faostat3.fao.org/browse/Q/QC/E, accessed on 8/16/2015). The five primary producers are China, India, the USA, Turkey and Egypt.

But Belgium, Netherlands, and Ireland are the producers with the highest yield. The yield in Belgium is up to 499,600Kg per hectare (http://faostat3.fao.org/browse/Q/QC/E, accessed on 8/16/2015). As shown in Table 1, an increase in tomato production is generally accompanied by an increase in the amount of land devoted to tomato cultivation. However, comparisons among the top ten tomato-producing nations suggest that there is still much room for improvement in increasing the crop's yield.

Table 1 The top ten tomato-producing countries, with estimated production, area harvested and yield in 2013 (source:FAO)

Country	Production(tonnes)	Area harvested(Ha)	Yield(Hg/Ha)
China, mainland	50,552,200	980,100	515,786
India	18,227,000	880,000	207,125
United States of America	12,574,550	149,977	838,432
Turkey	11,820,000	311,000	380,064
Egypt	8,533,803	212,946	400,750
Iran (Islamic Republic of)	6,174,182	163,595	377,407
Italy	4,932,463	95,304	517,550
Brazil	4,187,646	62,687	668,025
Spain	3,683,600	45,300	813,157
Mexico	3,282,583	87,165	376,594

Tomato production in the USA

Cultivated tomatoes are the third economically important vegetable crop after potatoes and lettuces in US in terms of per capita consumption (Foolad, 2007a). According to ERS data, U.S. tomato industry contributes more than $2 billion to farm income annually. The per capita use of selected, commercially produced, fresh and processing tomato in 2014 is up to 87.8 pounds. Due to its special flavor and high nutrient value, tomato is very popular in American daily life. Americans consumed 20.6 pounds of fresh tomatoes per person and 67.2 pounds of processed tomatoes in various ways: canned tomato, tomato pastes, tomato sauces etc. (http://www.ers.usda.gov/amber-waves/2015-september/potatoes-and-tomatoes-account-for-over-half-of-us-vegetable-availability.aspx#.Vj19qWDVxcU, accessed on 9/10/2015). In 2014, the nationwide production of tomatoes in the USA is 32.6 billion pounds in total, 8.5% of which are for fresh market and 91.5% of which are for processing market (Wells et al, 2014). The state of California is the largest processing tomato producer, with 96% share of market and the second largest fresh market tomato producer, producing 37% of the fresh tomatoes grown in the USA (Wells et al., 2014). Good climate (long, warm and drying growing seasons) makes California a natural suitable place to grow vegetables, especially tomatoes. Additionally, the use of hybrids, transplants and modern technology (e.g. laser leveling of fields, precision planting, advanced irrigation, fertilization) make California the bellwether of the whole vegetable industry in the US (Murray et al., 2001).

Usually, the processing tomatoes have several major traits, like determinate growth, dwarf plants, uniform fruit set pattern and tough skins, which are the characteristics for convenient mechanical harvesting (George, 1999). Besides, to make tomato paste more efficiently (Naeve, 2015), processing tomato varieties usually require a high level of soluble solids (averaging 5% to 9%).

In California, processing tomatoes are grown in rotation with other crops, such as garlic, onions, melons and wheat, in the San Joaquin Valley and Sacramento valley (Mitchell et al., 2001). Notably in 2015, water shortages in California did not inhibit processing tomato production. On August 28, the National Agricultural Statistics Service's California Processing Tomato report indicated the production for California processing tomatoes was 14.5 million tons, averaging 49.5 tons per acre. The current reported production exceeded 2014's amount by 4% and was the highest processing tomato production on record. The extended drought through recent years in California is anticipated to reduce the tomato production. Amazingly, the impact of drought has been mitigated by irrigation that supplies virtually 100 percent of California acres (Thornsbury & Jerardo, 2015). A majority of the operations were using conservative irrigation application methods such as drip, trickle, or low flow micro irrigation. However, a small number of California horticultural operations (40 operations) reported discontinued irrigation between 2012 and 2013 due to shortages of surface water or ground water (Thornsbury & Jerardo, 2015). If water supplies continue to tighten in the coming years, the increasing cost of irrigation will have significant effect on the price of tomatoes.

Generally, indeterminate tomato varieties for fresh production are grown in the greenhouse, while determinate varieties are grown in the field. In both types, tomato plants need to be transplanted, staked, pruned to increase fruit size and harvested by hand (George, 1999). There are plenty of varieties and cultivars for fresh production, ranging from small cherry tomatoes to large beefsteak tomatoes with various colors, shapes and flavors (Dorais, Papadopoulos & Gosselin, 2001).

Fresh-market tomatoes are grown in nearly every state, yet massive commercial production only exists in about 20 states. In 2014, more than 2.73 billion pounds of commercial fresh-market tomatoes were produced in the United States, valued at above 1.1 billion dollars (Wells et al., 2012). Florida and California are the two major fresh-tomato producing states, together accounting for over two-thirds of total U.S. fresh tomato acreage (Boriss & Brunke, 2005). Fresh tomatoes are planted in California in several counties during spring, summer and fall seasons (Boriss & Brunke, 2005). In Florida, tomatoes are planted intensively from November to January with harvest from October to June (Hochmuth, 2001). Most of Florida's winter tomatoes supply the eastern market in the nation (Girapunthong, VanSickle & Renwick, 2003). During the off-season period, Mexico and Canada exports fresh tomatoes to the U.S market (Girapunthong, VanSickle & Renwick, 2003).

However, in the US, up to 45% of land suffers from continuous or frequent drought stress, costing billions of dollars of losses in crops and business (Foolad, 2007a). Most crop plants, including tomatoes, are sensitive to drought stress throughout the whole life cycle. As the former biggest fresh-market producer (Florida is the current

biggest), California suffers from a continuous severe drought situation since 2011, resulting in nearly 12% decrease in fresh tomato production last year (2015), compared with average production from 2011-2013 (Thornsbury & Jerardo, 2015). Therefore, to guarantee fresh tomato yield in drought stress environments, drought tolerance is an urgent issue to solve now.

Tomato origin and distribution

The tomato and its wild ancestors originated from the Andean area of South America, which has a variety of habitats with climate ranging from arid to rainy (Lin et al., 2014). The most likely ancestor of cultivated tomatoes is *Solanum lycopersicum var. cerasiforme*, a kind of wild cherry tomato distributed in the South America (Rick and Holle, 1990). There are two competing hypotheses to identify the original place of domestication of tomatoes, one from Peru, and the other from Mexico (Robertson & Labate, 2007). Both hypotheses are suggested by linguistic evidence. In 1886, Alphonse De Candolle, the foremost authority on the origin of cultivated plants, suggested a Peru origin according to the names "mala peruviana" or "Pommi del Peru" named by botanists during the sixteenth century. Later in 1948, Jenkins advanced the Mexican origin hypothesis. This hypothesis came from a literature, in which it referred to tomato as "tumatle exTemixtitan", utilizing a native Mexican name (Jenkins, 1948). Intriguingly, he also argued "Pommi del Peru" was also used to refer to other *Solanaceous* plants, such as *Datura stramonium* L. and had no bearing on tomato, undermining De Candolle's linguistic proof.

Although the Mexican origin is reasonable, until now, we do not have explicit evidence to deny a Peruvian origin, or a parallel domestication in both areas (Robertson & Labate, 2007). The first history record of tomato is the description published in Old World by Andrea Mattioli of Italy in 1554 (Kallo, 1991). Then the Europeans brought tomato cultivars to the United States (Kallo, 1991). Thomas Jefferson described the first planting of tomato in Virginia in 1781, and 8 years later, French refugees took the tomato to Philadelphia in 1789, and in 1802, an Italian painter brought it to Massachusetts (Kallo, 1991). Since 1800, the tomato has been grown in most parts of the world and became a popular star in vegetables (Boswell, 1952).

Genetic diversity of tomato

Solanum section *Lycopersicon* includes the cultivated tomato *Solanum lycopersicum* and 12 additional wild relatives. Compared with most crop plants, the level of genetic diversity within cultivated tomato germplasm is very low due to domestication and human selection (Miller & Tanksley, 1990). Even using sensitive molecular markers, very few polymorphisms have been identified within tomato cultivars (Bai & Lindhout, 2007). The lack of diversity within tomatoes is not a barrier to breeding progress, because the variation within the genus *Solanum* is tremendous (Heuvelink, 2005). Recently, breeding efforts have enriched genetic bases of tomato cultivars with novel alleles that improve productivity and adaptation via introgression with wild relatives (Liedl et al. , 2012).

Variation in levels of genetic diversity within and among population of a species can result from mating system and human selection (Liedl et al. , 2012). Mating system

is one of the factors that have been thoroughly studied for the genetic diversity of *Lycopersicon*. Mating system has also played a significant part in wild tomato species evolution, varying from allogamous self-incompatible, to facultative allogamous, to autogamous and self-compatible (Bauchet & Causse, 2010). As a consequence of their self-incompatibility and outcrossing mode of reproduction, the outcrossing species may have greater genetic variation than the largely self-pollinating species (Bauchet & Causse, 2010). Compared to random mating, selfing is anticipated to reduce genetic variations, by the decrease of effective population size, frequent bottlenecks and reduced recombination (Robertson & Labate, 2007).

Although all wild tomato species are diploid (2n = 2x = 24) and theoretically can be hybridized with the cultivated tomato, difficulties arise when using special techniques (Stevens & Rick, 1986). Stevens & Rick (1986) recognized two species complexes according to the ability of hybridization with cultivated tomato: the *Esculentum* complex (*S. esculentum, S. pimpinellifolium, S. cheesmaniae, S. pennellii, S. hirsutum, S. chmielewskii and S. parviflorum*), which was intercrossable with higher success rate, and the *Peruvianum* complex (*S. chilense and S. peruvianum*), which had poor hybridization capacity (Foolad, 2007a).

Until now, most of the wild tomato species have been extensively utilized for improving the disease resistance in the tomato crop. In addition, they have been used as a source for resistance to parasitic plants, tolerance of abiotic stresses and enrichment of quality traits (Robertson & Labate, 2007).

Some accessions of the *S. hirsutum* have been characterized as near-completely resistant to early blight caused by the fungus *Alternaria solani* in the field and greenhouse experiments (Foolad, Ntahimpera, Christ, & Lin, 2000). High levels of resistance to tomato yellow leaf curl disease (TYLCD), one of the most devastating diseases in the last few decades, were found in several wild species, including *S. pimpinellifolium, S. peruvianum, S. chilense, S. habrochaites and S. cheesmaniae* (Ji et al. 2007). Three resistance QTLs were identified for grey mold resistance in an F2 population and confirmed in BC_2S_1 population deriving from *S. lycopersicum* cv. Moneymaker x *S. habrochaites* LYC4 cross (Finkers et al. 2007).

Several studies have identified QTLs in wild species for improvement of some horticultural traits. For instance, *S. hirsutum* alleles were detected that gave 16% increase in total yield and *S. pennelli* achieved 20% improvement via introgressions (Robertson & Labate, 2007). Bernacchi et al. (1998) reported genetic gains had been fulfilled by hybridization with *S. hirsutum* and *S. pimpinellifolium* for desirable QTLs controlling quality traits like fruit firmness, and soluble solids contents.

Important breeding traits

Depending upon location, market need, climate, time and resources, the goals of tomato breeding programs are set specifically. Generally in history, goals of tomato breeding have undergone four phases: breeding for yield in the 1970s, for shelf-life in the 1980s, for taste in the 1990s and currently for nutritional quality (Bai & Lindhout, 2007). Besides, the ongoing challenged of biotic and abiotic stresses are always high priorities for breeding programs, since these factors affect the tomato yield. The tomato

seeds, which growers want to buy, must have the potential to produce a high-yield and high-quality fruit, still keeping production costs as low as possible.

The bottom line of breeding in any crop is to guarantee the production of an economically important part of the plant (Panguluri et al., 2013). Similarly higher yield is also the major objective for both fresh market and processing tomato breeding. As mentioned before, production of tomatoes in the world increased four-fold in the last 50 years. This boost in crop productivity is due to increased plant area, improved cultivation practices and plant breeding (Panguluri et al., 2013).

Heterosis in tomato can be commercially harnessed considering the advantages in the earliness, uniformity, total yield and resistance attributes (Kallo, 1991). The first commercial tomato hybrid cultivar 'single cross' was released in 1946 (Bai & Lindhout, 2007). Currently, tomato cultivars for the fresh market are hybrids. Meantime, there is an accelerating trend using hybrid seed in the processing market (Bai & Lindhout, 2007). If breeders can create hybrid vigor in tomato yield, this would be good news for the industry. Recently, a research reported that a single overdominant gene, single flower truss (SFT), drove tomato heterosis in yield, by increasing total yield up to 60% in a heterozygous mutant (Krieger, Lippman, & Zamir, 2010). This heterosis was stable, repeatedly occurring in diverse genetic backgrounds and environments (Krieger, Lippman, & Zamir, 2010). The authors suggested using mutants might be a new strategy to identify genes with heterosis and provide innovative germplasm for tomato breeding.

Breeding for resistance to most disastrous pests and pathogens is another prominent goals in tomato breeding. Many of resistances are simply inherited, and remarkable successes have been demonstrated through backcross introgression of genes from wild species. Potential for creating disease resistant mutants using transgenic tools has also been proven (Bai & Lindhout, 2007). One of the first examples was the exploitation of *Cladosporium fulvum* (tomato leaf mold) resistance from *S. pimpinellifolium* in 1934 (Bai & Lindhout, 2007). Although tomato can grow in the most areas in the world, unfavorable growing environments, such as drought, excessive salinity, or high temperature, have caused about two-thirds of tomato yield potential to routinely be lost (Cortina & Culiáñez-Macià, 2005). Many genetic variations for such stress tolerances exist in wild species and germplasms of tomatoes, which can be exploited in future breeding projects.

GBS-SNP markers

Plant breeders select traits of interest usually by observing phenotypes and then make the selections. However, when the selected traits are not easily to observe, such as the underground part or the invisible physiological traits, it could be challenging and time-consuming to accurately phenotyping them. With the development of molecular biology technology, especially the next generation sequencing technology, molecular markers enable breeders to select desired traits in the early stage of plants or even without actually planting the crops in the field, which takes time and cost a lot of space. Since environment may have significant effects on some traits, if only based on the phenotypic selection, we may not be able to select genetically heritable trait.

Among the DNA-based molecular markers, SNP (single nucleotide polymorphism) becomes popular for its high abundance in the genome and simplicity by just following a bi-allelic genetic mode. SNP markers are the locus specific markers, which can be scored co-dominantly in a flexible way (Víquez-Zamora et al., 2013). Genotyping by sequencing (GBS) is a cost-effective way to discover genome-wide SNP markers for population, by its simple library preparation and combination the power of multiplexed NGS with enzyme based genome reduction (Spindel et al., 2013). It has been estimated that GBS was capable to generate thousands of SNP markers for a large population at an expense about $9 per sample for 384-plex (Spindel et al., 2013). With these advantages, GBS becomes an important and powerful tool for yielding SNP markers for genetic mapping. By choosing an appropriate restriction enzyme, repetitive sequence regions of genomes can be avoided, which will greatly reduce the complexity of challenging alignment problem in the final bioinformatics analysis (He et al., 2014). The genome size of tomato is about 950 Mb, composed of pericentromeric heterochromatin and distal euchromatin (Tomato Genome Consortium, 2012). Pericentrometic heterochromatin, rich in repetitive sequences, is estimated about 75% of tomato region (Garcia-mas, 2016). A study characterized *S. lycopersiucm* genome consisting of 50~60% repetitive elements (Mehra, Gangwar, & Shankar, 2015). Applying the GBS for SNP discovery in tomato species, by using the restriction enzymes to filter out the repetitive sequence region and reduce the genome sequencing length can reduce the cost of genotyping, and at the same time does not need to compromise the efficiency to detect the SNP markers. Also, with the completion of

tomato reference genome sequence, localizing SNPs generated by GBS to specific physical locations is much easier than before (Foolad & Panthee, 2012).

Linkage map and QTL analysis in tomato

Constructing a linkage map by genetic markers or known genes is based on their recombination frequency during the crossover. Genes on the different chromosomes segregate independently and produce 50% recombinant gametes. Genes on the same chromosome are linked and only produce parental gametes. However, at the first phase of meiosis, sometimes the two non-sister chromatids exchange partial segments. This process can break up the linked genes and produce the recombinant gametes. The probability of a crossover causing recombination between genes is less than 50%. Sturtevant, a student of Morgan's, proposed the greater the distance between linked genes, the greater the chance the crossover could happen. According to this relationship he created the first genetic map of fruit fly.

From then, many tomato genetic maps have been generated through the years. As explained earlier, there is little genetic diversity among tomato cultivars. So most of the tomato genetic maps were among interspecific crosses. *S. pennelli* as the most distant from the *S. lycopersicum,* was used extensively for genetic mapping. Recently, there were two high-density linkage maps between *S. lycopersicum* × *S. pennelli*, EXPEN 2000 and EXPEN 2012 maps developed, consisting of 3503 and 3687 markers, representing 1076 and 1229 unique map positions, respectively (Sim et al., 2012). Another wild species *S. pimpinellifolium* is also popular for mapping experiments, because of its close relationship to *S. lycopersicum*, ease of crosses and red-fruit

characteristics (Foolad, 2007b). An genetic map was generated with 4491 markers (1358 bins) and average bin interval of 0.8 cM (Sim et al., 2012). These three high-density genetic maps revealed very good collinearity between the genetic map and the physical map except very small regions on chromosome 3, 10 and 12 (Chen et al., 2014).

Using linkage maps to facilitate QTLs analysis for complex traits in tomato has been around for three decades. Many QTLs have been identified for fruit traits and disease resistant traits and some of them are successfully applied for marker-assisted breeding. In the tomato seed industry, DNA markers which are linked to the QTLs, are routinely employed for selecting for disease resistance traits, including fusarum wilt races 1, 3, late blight (Ph-3) and tomato yellow leaf curl (Ty1, Ty2, Ty3 and Ty4) (Foolad & Panthee, 2012).

Root genetics and breeding in tomato

The root is an essential organ with various functions to support plant growth, such as absorbing water and nutrients, anchoring in the soil and interacting with the microbial community in the rhizosphere (Kuijken,Van Eeuwijk, Marcelis & Bouwmeester, 2015). Modifying root traits in plant breeding program can result in more stress-tolerant crops and possibly increase yield by enhancing the capacity of plant for water and nutrition acquisition(Paez-Garcia et al., 2015). However, like other vegetable breeding programs, tomato breeding programs mainly focus on improving disease resistance and above-ground traits, while root traits are overlooked, probably due to the difficulty of phenotyping the underground parts of plants, and lack of knowledge about the genetic control of root development.

As we know, screening root architectural traits often requires destructive approaches, which may reduce the accuracy of the measurements. When separating the whole root system from soil, it is hard to avoid damage to the roots. Also, it takes a huge amount of time and labor to analyze root parameters, such as total root length, etc. Moreover, most of the analysis is done at the end of the experiment and thus monitoring root system dynamic growth over a plant's life cycle is not feasible (Shabala, 2012). Now several encouraging methods emerge, using root observation chambers, soil-less media or image-based and 3D phenotyping platforms to analyze root parameters (Shabala, 2012). These techniques provide opportunities to understand the root traits and are applied in root breeding research.

Tomato has been treated as a model plant in many gene expression researches to study the single gene controlling relevant root trait, although most of genes are for root-knot resistance and physiological-biochemical response. After searching the QTLs for root traits in the Sol Genomics database, until now there are few QTLs and markers reported for selecting root system architecture. The QTLs study for root system architecture in other agronomic crop has gained a lot attention. One successful example is that a QTL for root depth on chromosome 9 of rice, which has been deployed in marker-assisted selection, has shown its great effectiveness to increase the yield in water-limiting conditions(Uga et al., 2013). Developing tomato cultivar with efficient root system is necessary for increase the yield and better resistance to the abiotic stresses, especially drought, and root diseases.

Research objective

The main purpose of this study was to discover SNP markers for constructing a genetic linkage map based on an F2 population derived from a cross between *S. lycopersicum* and *S. cheesmaniae* and to use the genetic map to identify the position of QTLs for root and other morphological characteristics.

This study is designed to answer the following questions:

1. Is it feasible to select and modify root traits through plant breeding? What is the genetic control behind them?

2. What is the relationship among the root traits and other morphological traits when selecting them?

3.Are there any QTLs associated with the root and other morphological traits? Can we develop SNP markers to help us to select desirable root traits, because root phenotyping is difficult and time consuming?

CHAPTER II

GENETIC VARIATION IN TOMATO ROOT

The very first step to start a QTL discovery project is to create a population with traits of interests segregating. The genotypes with opposite traits need to be selected and then to create a mapping population. In our study, two lines with different root traits were selected for this study after a preliminary investigation on several root and morphological traits in February of 2015. The objective of the initial experiment is to estimate the phenotypic variation, heritability and correlation among root traits and other morphological traits. The null hypothesis is that the root traits are simply inherited and are the same in the different genotypes.

Materials and methods

Plant materials

Two phenotypically different parents were selected to create the F2 segregated population. One was a wild tomato species Lche4 (*Solanum cheesmaniae*), from arid Galapagos Island, and the other one was RvT1 (*Solanum lycopersicum*), a TAMU advance breeding line with favorable horticultural traits.

There are several reasons why we use *Solanum cheesmaniae* as the donor parent. Firstly, it is easy to cross within tomato cultivars and have fertile progenies; secondly, it can survive in extremely arid and saline environments, which may also present high tolerance to salinity and drought; thirdly, it contains high sugar and beta-carotene

contents, which is also a good source for nutrition improvement in tomato breeding; fourthly, it exhibits resistance to whiteflies(Firdaus et al., 2013; Lanca, 2004). A preliminary study of the root traits between the two parents was also conducted in Feb 2015, which showed Lche4 had a significant small root system compared to RvT1.

As shown in Figure 1, RvT1 as the female parent was crossed to Lche4. The F1 was self-pollinated to produce an F2 population. As well, a commercial cultivar Hot-Ty was used as the environmental control.

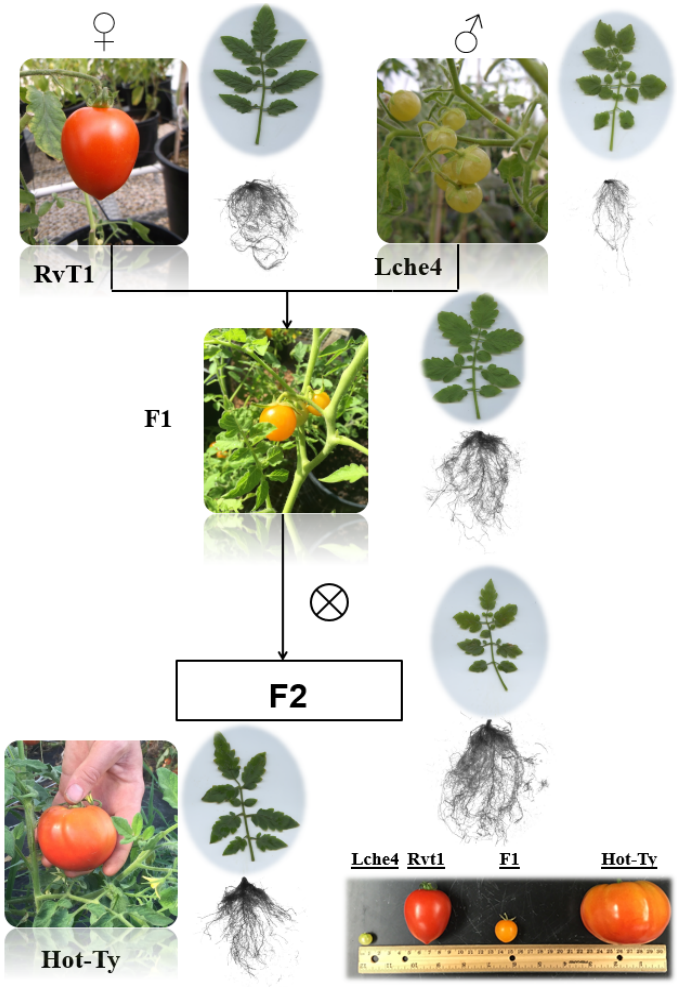

Figure 1 General morphology of parents, F1 and Hot-ty.

Measurement of root and other morphological traits

Seeds of the F2 population, both parents and environment control were planted in a completely randomized design. On November 11th, 2015, all the plants were planted in six 38-cell plant trays, with volume 220.08 mL per cell, filled with sand media in the greenhouse at Texas A&M University (College Station, TX). At the bottom of the cell, two cotton balls were placed to prevent sand falling out. Six trays were put on the heat

mat, keeping the soil warm and helping germination in the cold winter. After germination, the heat mat was removed. Initially, two seeds per cell were sown while only a single seedling was allowed to grow when seedlings were established. 1 tablespoon of Peters professional 20-20-20 soluble fertilizer mixed in one-gallon of water was applied every week. Spray application of pesticide was used to keep the plants and greenhouses free from whiteflies and powdery mildew. However, the greenhouse was still affected by whiteflies.

Twelve weeks after planting, 181 F2 seedlings as well as 5 RvT1, 5 Lche4, and 10 'Hot-Ty' were harvested before flowering. The content of chlorophyll was measured independently on three fully expanded leaves using Minolta SPAD 502 Plus Chlorophyll Meter. The average was used as the final SPAD value. The shoots of all the plants were cut, and the shoot length, fresh weight of shoot and stem diameters were measured. Then the shoots were treated with rooting hormone (Green Light®, Rooting Hormone), and placed in vermiculite media under the mist bench. These genotypes were later used to produce the F3 generation. The roots were extracted by putting the entire pot on a fine wire mesh sieve and immersing into a bucket filled with water. All the sand rapidly flowed through the holes of the sieve by gently shaking the sieve, while all roots remained in the sieve. The attached cottons were gently removed by hand and the roots were put into Ziploc bags. These roots were stored in a refrigerator at 4 ^{o}C for up to 5 days while the root analyses were completed.

Each root was placed on a transparent plate with a shallow film of water in order to separate the lateral roots and to minimize the overlap area. The roots were scanned as

tiff format pictures by an Epson scanner, with a transmitted light source to avoid shadows. After scanning, the entire root system was placed on the paper towel for 30 minutes to remove excess moisture and than put on a digital balance to record fresh weight. The roots were put in envelopes and dried in an oven at 80 ^{0}C for 72 hours before the dry weights of roots were measured. The batch root analysis was performed in WinRHIZO root-scanning software (Version 2013a; Regent Instruments Inc., Ottawa, ON Canada) using the scanned pictures of the roots. The total root length (TRL), root surface area (RSA), root volume (RV) and root average diameter (RAD) were acquired directly in the software. Root to shoot ratio was defined as root fresh weight/shoot fresh weight (R: S). The ratio of root dry weight to root volume was computed as root tissue density (RTD). The specific root length (SRL) was calculated as total root length divided by root dry weight.

Statistical analysis

All the data were analyzed by JMP software (Version 12.0; SAS Institute Inc. 2015). Variance (ANOVA), least significant difference Student's t test, normality test and Pearson's correlation coefficients were obtained. The ANOVA and Fisher least mean test were performed among the two parents and the environmental control. The Shapiro-Wilk test and Pearson's correlation coefficients were calculated among 14 traits in the F2 population for normality and correlation tests.

Results and discussion

Traits variation among parents and environment control

The 13 traits showed significant differences among genotypes according to the ANOVA test. Each trait mean was separated by the LSD test. As shown in Table 2, all the root and morphological traits were highly different between parents, except the shoot length and root tissue density. The female parent RvT1 had a much higher value than the male parent Lche4 of RFW, RDW, RL, RS, RV, RAD, R:S. However, interestingly, the specific root length of Lche4 was 2 times more than that of RvT1. Generally, the measurements indicated that RvT1 had a much larger root system than Lche4. The only exception was the specific root length with a significant decrease. Probably, it is because Lche4 was a wild species, which lives in the dry environment. Many species of unproductive, nutrient-limited or dry environments have been observed to have an increase in SRL, compared to the species of productive environments (Ryser, 2006). One explanation for high SRL in plants is a relatively small initial investment in biomass per unit root length, which could help them to exploit pockets of water or nutrients in the soil (Eissenstat, 1991). Many reports indicate that high SRL and greater root length density enabled plants to extract water more rapidly than those with low SRL (Eissenstat, 1991). Higher specific root length with a relatively big root system can possibly be a desirable trait to select when breeding for drought resistance.

Table 2 Mean separations among parents and environmental control

Genotype	RvT1	Lche4	Hot-Ty
SL(cm)	17.08±1.96 b	14.96±1.72 b	23.31±2.76 a
SFW(g)	12.72±2.65 b	6.02±1.07 c	16.31±2.39 a
SD(mm)	4.56±0.80 a	3.51±0.60 b	4.92±0.59 a
SPAD	39.30±1.31 b	18.74±1.83 c	44.96±2.91a
RFW(g)	3.75±0.78 b	0.47±0.19 c	5.06±0.88 a
RDW(g)	0.28±0.06 b	0.04±0.01 c	0.43±0.06 a
R:S	0.30±0.04 a	0.08±0.04 b	0.32±0.08 a
RL(cm)	1467.60±244.07 a	467.01±155.20 b	1881.14±501.24 a
RSA(cm^2)	246.57±44.73 b	56.70±18.11 c	329.77±70.11 a
RV(cm^3)	3.31±0.71 b	0.56±0.21 c	4.65±0.96 a
RAD(mm)	0.53±0.04 a	0.39±0.06 b	0.57±0.04 a
SRL(m/g)	52.92±9.22 b	107.20 ±27.86 a	43.07±8.90 b
RTD(g/cm^3)	0.09±0.01 a	0.08±0.02 a	0.10±0.01 a

*All the trait means were separated by LSD Student's t test at α=0.05 level. The same letter indicated there was no significant difference between the two values.

The correlation among traits

As show in Figure 2, the blue color showed positive correlation while the red color showed negative correlation. When the color changed from light to dark, the Pearson correlation coefficient changed from small to big. The correlation coefficients at a 0.05 significance level were shown in the matrix. A combined scatter, histogram and correlation plot with detailed information was in the Appendix.

The Pearson correlation coefficient is to measure of the strength of a linear relationship between two variables. However, if the relationship is not linear, the correlation coefficient will not adequately represent the strength of the relationship. For example, when $y = x^2$, the correlation coefficient equals to 0, even though y is determined by x. So we cannot say y is not correlated with x in this situation. In our study, there were three calculated root traits: specific root length, root tissue density and root to shoot ratio. Take specific root length as an example, it was determined by root length and root dry weight. The correlation coefficient between specific root length and root length was 0.31. The correlation coefficient between specific root length and root dry weight was -0.16. The two correlation coefficients were both relatively low. Because the relationship between specific root length with these two variables was not linear, some higher-order relationships may exist. After testing with a mixed model ($z=ax+by+cxy$) among specific root length, root length and root dry weight, the correlation coefficient was up to 0.86. In the following discussion, we only consider the linear relationship between two variables. When we say two traits are correlated, we indicate they have a linear correlation. When we say they are not correlated, it will have two meanings: they are independent of each other or they may have a non-linear relationship.

One interesting finding was that the specific root length negatively correlated with root average diameter, and root tissue density with the coefficients of -0.64, and -0.61, respectively. Similar results were obtained in a citrus root study, in which high specific root length correlated with lower tissue density (Eissenstat, 1991). The root length, root surface area and root volume had very high positive correlation coefficients

with each other, with the values more than 0.95. The root-to-shoot ratio was also highly correlated with root length, root surface area, root volume, root fresh weight and root dry weight, with correlation coefficients more than 0.70. The shoot fresh weight was moderately correlated with root length, root surface area, root volume, root fresh weight and root dry weight. There was no correlation between root length and root average diameter.

It is important to identify positive and negative correlations among traits, since selection for one trait may impact selection for other desirable traits. In our study, selecting higher specific root length meant compromising with lower values in any other root parameters. However, for the shoot fresh weight, root length, root surface area, root volume, root-to-shoot ratio, root fresh weight and root dry weight with positive coefficients, they can be simultaneously selected. Selecting for root average diameter may be independent from selecting the root length.

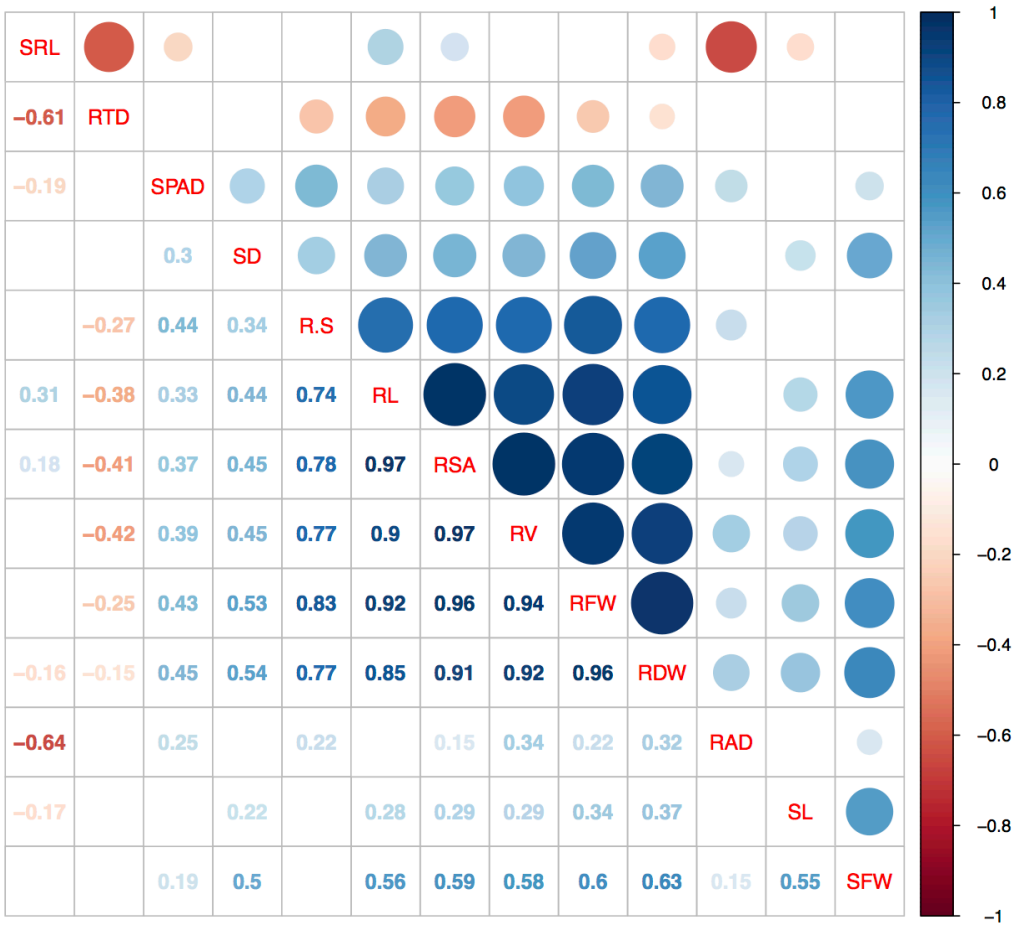

Figure 2 Correlation coefficients matrix among 13 traits in the F2 population from RvT1 x Lche4

The frequency distribution of traits

The frequency distribution histograms of 13 traits are shown in Figure 5. All the traits were fully and continuously segregated, which suggested polygenic control behind them. The null hypothesis that a single gene or two genes controlled the traits was rejected, because the distributions didn't show this simple pattern. After a Shapiro-Wilk

test, all the shoot traits, including shoot length, shoot fresh weight, shoot diameter and SPAD, were following a normal distribution, which would satisfy the assumption for common QTL detection methods. Most of the root traits showed a right skewed distribution in the F2 population, except for root average diameter, which passed the normality test. These skewed traits might need to be transformed to do the QTL analysis or a nonparametric method might be required.

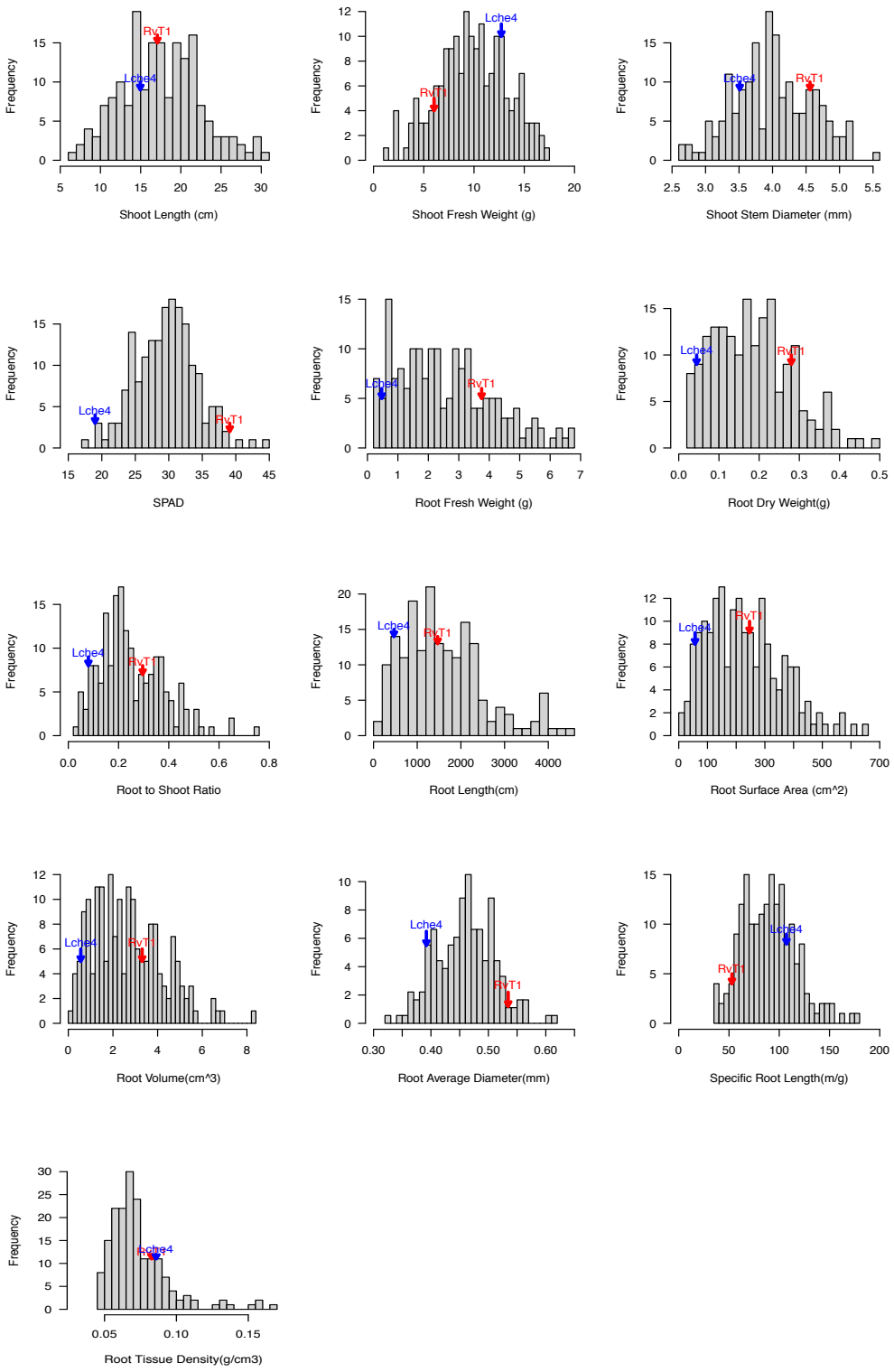

Figure 3 Frequency distribution of 13 traits among the F2 population from RvT1x Lche4

Table 3 Shapiro-Wilk test for normality

	SL	SFW	SD	SPAD	RFW	RDW	R:S	RL	RSA	RV	RAD	SRL	RTD
P-value	0.27	0.28	0.59	0.89	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.73	0.02	<0.01

The inheritance of selected traits

Broad sense heritability for selected traits was estimated as the ratio of genotypic variance to phenotypic variance, expressing the extent to which an individual's phenotype was determined by genotypic factors (Napier, 2006). The environment control "Hot-Ty" had genetically uniform plants, so the variance of selected traits for Hot-Ty, was assumed to equal the environmental variance (VE). The variance of the F2 population was assumed to be the phenotypic variance. By subtracting VE from VP, the VG (Genetic variance) was determined (Napier, 2006).

As shown in the Table 4, the broad sense heritability of shoot length, shoot fresh weight and SPAD were more than 50%, implying these traits were moderately heritable. However, the heritability of stem diameter was negative. This could suggest that the trait was largely affected by environmental factors. Among the root traits, the heritability of specific root length was 89.17%, which suggested a large genetic contribution to this trait. The root length, root surface area, root volume, root dry weight, root fresh weight and root tissue density also had relatively high heritability values around 70%, while the root average diameter was lowly heritable traits, with values of 28.03%, respectively.

The higher heritability values hinted at a better chance of reliably detecting QTLs with large effects (Crespi, 2013). The results indicated a good possibility to detect QTL for one or several traits.

Table 4 Phenotypic, genotypic and environmental variance and corresponding broad-sense heritability for 13 traits in the F2 population from RvT1 x Lche4

Trait	SL	SFW	SD	SPAD	RFW	RDW	R:S
VP	24.0133	11.8721	0.3417	20.4422	2.4737	0.0103	0.0166
VG	16.3812	6.1759	-0.0044	11.9929	1.6927	0.0064	0.0099
VE	7.6321	5.6962	0.3461	8.4493	0.7809	0.0039	0.0067
H	0.6822	0.5202	-0.0129	0.5867	0.6843	0.6181	0.5950

Trait	RL	RS	RAD	RV	SRL	RTD
VP	885646.5036	17277.7709	0.0026	2.4532	729.7498	0.0004
VG	634409.1185	12362.9830	0.0007	1.5299	650.6904	0.0003
VE	251237.3851	4914.7879	0.0019	0.9234	79.0594	0.0001
H	0.7163	0.7155	0.2803	0.6236	0.8917	0.6995

CHAPTER III

SNP DISCOVERY AND LINKAGE MAP CONSTRUCTION IN THE RVT1 × LCHE4 POPULATION

As there already exist high-density genetic maps in tomato, why should we continuously develop new genetic maps? There are several reasons: first, not all genes of interest will segregate in the selected mapping population; second, the markers developed in the map are not always polymorphic and useful in other mapping populations (Ashrafi, 2007). Due to these reasons, it is necessary to develop new markers to construct new genetic maps, in order to exploit the genetic potential of wild species.

Several linkage maps between *S. lycopersicum* and *S. cheesmaniae* have been reported. The first linkage map between the two species was created with 71 RFLP markers in a 350 plant F2 population (Paterson et al., 1991). Another linkage map was constructed with 132 RFLP markers among a 97 RIL population (Paran, Goldman, Tanksley, & Zamir, 1995). (Villalta, Reina-Sánchez, Cuartero, Carbonell, & Asins, 2005) used 114 SSR and SCAR markers to construct a genetic map among 115 RILs.

These three maps are low-density linkage maps, as thousands of SNP can be easily generated today using next generation sequencing technology. There was no linkage map reported between these two species by using SNP markers. By applying the GBS method, in this research, a relatively high-density genetic map using SNP markers

was generated. The genetic map with dense markers increases our ability to detect QTLs and further, to search for candidate genes.

Materials and methods

Genotyping by sequencing

Very young leaf tissues of 181 F2 plants along with the parents were sampled in 1.5 ml tubes with grinding beads and immediately stored in liquid nitrogen. 50-100 mg of young leaf tissue was collected from each individual plant and placed in a 1.2 mL tube with two grinding ceramic beads. All the tubes were stored in an -80 °C freezer for later genotyping work. Leaf sampling was carried out three times in case of failure in the DNA extraction. Considering that tomato leaves contain lots of polyphenols, which could contaminate the DNA sample during the extraction process, we used a modified microprep extraction protocol (Fulton, Chunwongse, & Tanksley, 1995). The fresh microprep buffer was prepared and kept well mixed at room temperature and then warmed in a 65 ^{o}C water bath. 450 µL of microprep buffer containing 2.5 parts DNA extraction buffer (0.35 M sorbitol, 0.1 M tris-base, 5 mM EDTA, pH 7.5), 2.5 parts nuclei lysis buffer (0.02 M tris, 0.05 M EDTA, 2 M NaCl, 2% CTAB) and 1.0 part 5% Sarkosyl were added to each tube. The samples were quickly ground into a liquid form using FastPrep-24 homogenizer with the Tomato, Early girl, Leaf tissue program.

After being incubated in a 65 °C water bath for 90 minutes, the tubes cooled down for 5 minutes to room temperature. 500 µL Chloroform: IsoAmyl alcohol (24:1) solution was added into each tube. After making sure the lids were closed tightly, a paper towel was placed between the tube lids and the plate lid to absorb any liquid that

might have spilt. Pressing on the lids, the tubes were shaken and inverted for 2 minutes, and later were centrifuged for 15 minutes at 4000 rpm. The resulting supernatant liquid was transferred into a new 1.2 mL deep well tube. 500 µL of cold Isopropanol was added to each tube, mixed gently, and put in the freezer at -20 ^{o}C for overnight. The tubes were later centrifuged for 15 minutes at 4000 rpm again. After the isopropanol was poured off, the DNA pellet was attached to the bottom of the tube. The pellet was washed with 100 ul 70% Ethanol, and 400 ul 70% Ethanol was added after the wash. Tubes were centrifuged for 5 minutes at 14000 rpm and the supernatant was removed. The tubes were left open on a bench for 15 minutes to allow the pellet to dry completely. At last, the DNA pellet was re-suspended in 90 ul of TE buffer.

The Zymo DNA clean and concentrator kit was used to purify the stock DNA. The purified DNA was quantified by a Qubit fluorometer (Invitrogen) using the dsDNA BR assay kit. The DNA content of each sample was measured and recorded. 500 ng aliquots of DNA from each sample were arranged in a 96-well plate and the total volume of each well was increased with dH_2O up to 16.8 ul. Later Mrs. Natalie Patterson prepared the GBS library. The library template was sent to TAMU sequencing center for running on the Hiseq-2500 machine. The sequence alignment and SNP calling was done by Dr. Patricia Klein.

F2 genotype coding

The discovered SNP markers in the population were coded by the following rule. The SNP allele of F2s, which is homozygous to the female parent RvT1 was coded as A and male parent Lche4 was coded as B. An H code was given to the F2s with

heterozygous SNP genotypes. The missing allele and bad allele calls were coded as dashes.

SNP marker data checklist

A perfect set of marker data for constructing a linkage map requires no missing data, no genotyping error and no segregation distortion. However, real world data are usually affected by these factors (Hackett & Broadfoot, 2003). Before we start to draw a linkage map, it is necessary to have a basic sense about the marker data and make a strategy to choose a subset of high-quality markers. The low-quality and problematic markers would increase the expansion of the map if included. The R package "R/qtl" and "ASMap" offered diagnostic functions to visualize and check potential mistakes for the genotypic data (Broman & Sen, 2009; Taylor & Butler, 2015).

The checklist for an unconstructed marker set included missing rate, genotype frequency, biological clones, the number of crossovers and double crossovers, which were checked for every individual. These factors could affect the linkage map quality. High rates of missing data might imply a genotyping problem. Abnormal genotype frequencies might reveal the individuals with high genotyping error or even some other labeling mistake (Broman, 2001). Excessive crossovers and double crossovers reflect departures from Mendelian genetics (Taylor & Butler, 2015). The statGen() function was used to examine the above factors. Sampling error presumably could be a result of biological clones or the individuals sharing a high proportion of alleles with each other. This kind of duplication happens quite infrequently (Broman & Sen, 2009). The R function clonestat() was used to detect similar individual pairs.

For every marker, segregation distortion, duplicity and number of double crossovers were checked. Any significant segregation distortion would reduce the map accuracy and further diminish the statistical power to detect the QTL. The segregation distortion was checked by a χ^2 test of the Mendelian proportion. The p-value used was after Bonferroni correction (p-value/number of markers) for the multiple tests (Broman & Sen, 2009).

Markers with moderate departure from Mendelian frequency are not rare. This distortion could occur, if some segments of chromosomes have lethal genes (Broman & Sen, 2009). Genotyping error is usually exhibited in the form of tight double crossovers. Unexpected double crossovers could suggest excessive genotyping error. The R function profileMark() was applied for checking these three factors among every marker.

Construction of a linkage map

The selected markers were imported into the JoinMap software (version 4.1,Kyazma©). Using independent LOD score of 10, linkage groups were correctly assigned to chromosomes according to the known locations of SNP markers. Because the tomato reference genome was well built and the reported tomato genetic maps have shown collinearity with physical maps, the marker order was fixed based on the physical position using fix order functions. The LOD score ≥ 6, recombination frequency ≤ 0.4, goodness-of fit jump threshold =5, ripple value=1 were set as parameters to build the linkage map by regression mapping algorithm. The Kosambi function was applied to convert the recombinant frequency to map distance.

Results and discussion

Genome-wide SNP markers

After aligning to the tomato reference genome, 1002 polymorphic SNPs were found, excluding the bad markers and markers with more than 15% missing data. As showed in Figure 4, most of the markers were located on the upper and lower arms of the chromosomes, with less in the center region. The tomato genome consists of 50~60% repetitive elements in the pericentromeric heterochromatin, and the restriction enzyme NgoMIV didn't cut these regions. Thus the markers were mainly distributed on the chromosome arm, which has been seen in similar tomato genome research.

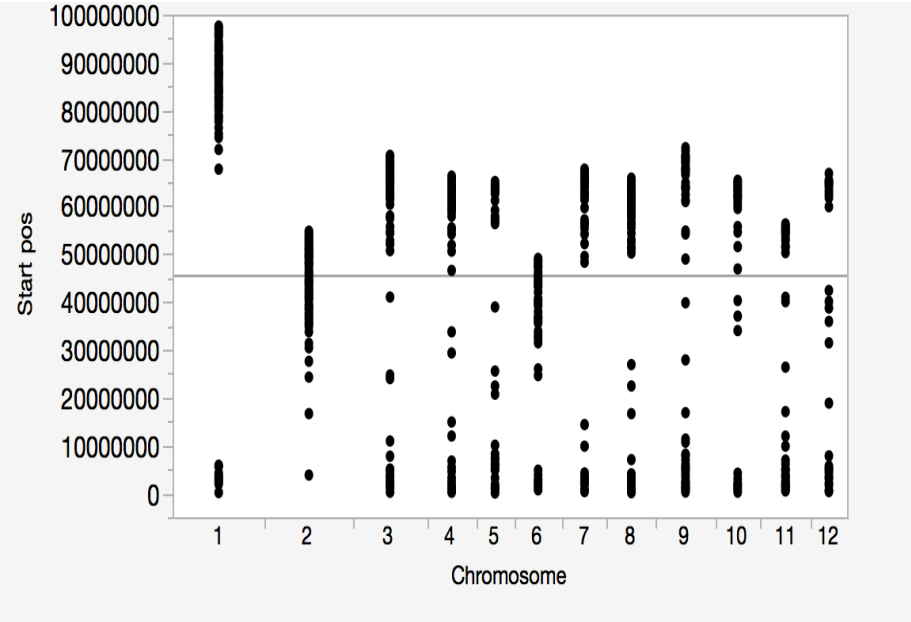

Figure 4 Physical position of SNP marker on the chromosome

SNP marker data checking

As shown in Figure 5, the AA and BB genotypes' frequencies among individuals are around 0.2~0.4. The AB heterozygous genotype frequency among individuals was around 0.4~0.6. The AA, AB, BB genotype frequencies among the F2 population are expected as 0.25, 0.50, 0.25, respectively. The results didn't show that odd individuals existed, such as one with ~100% AA or BB genotypes. The genotypic data was good in this aspect.

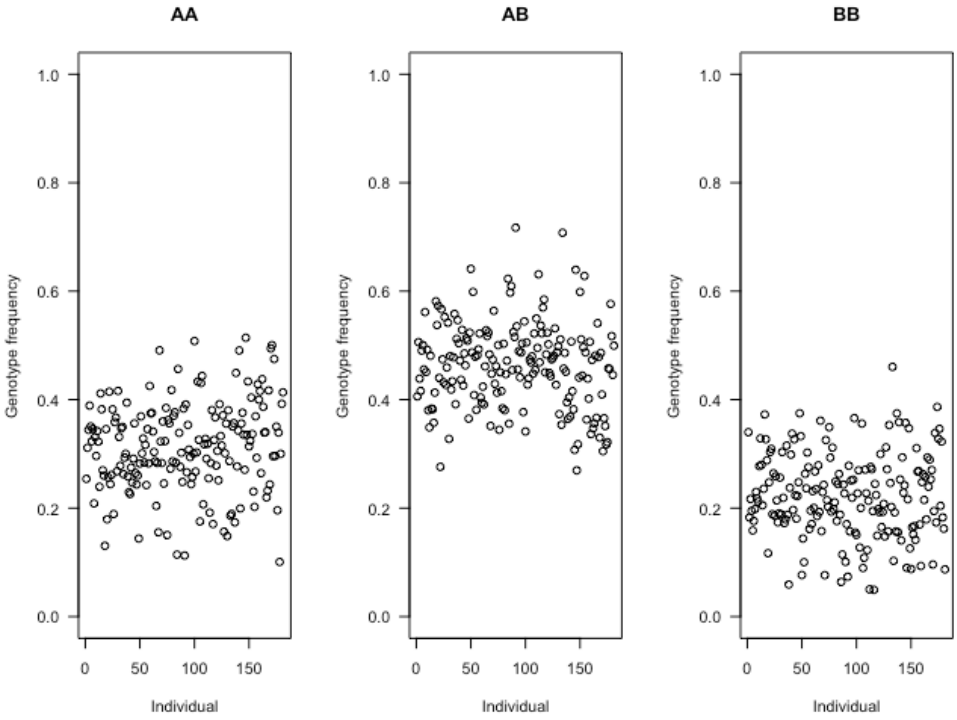

Figure 5 Genotype frequencies among individuals consisting of the F2 population

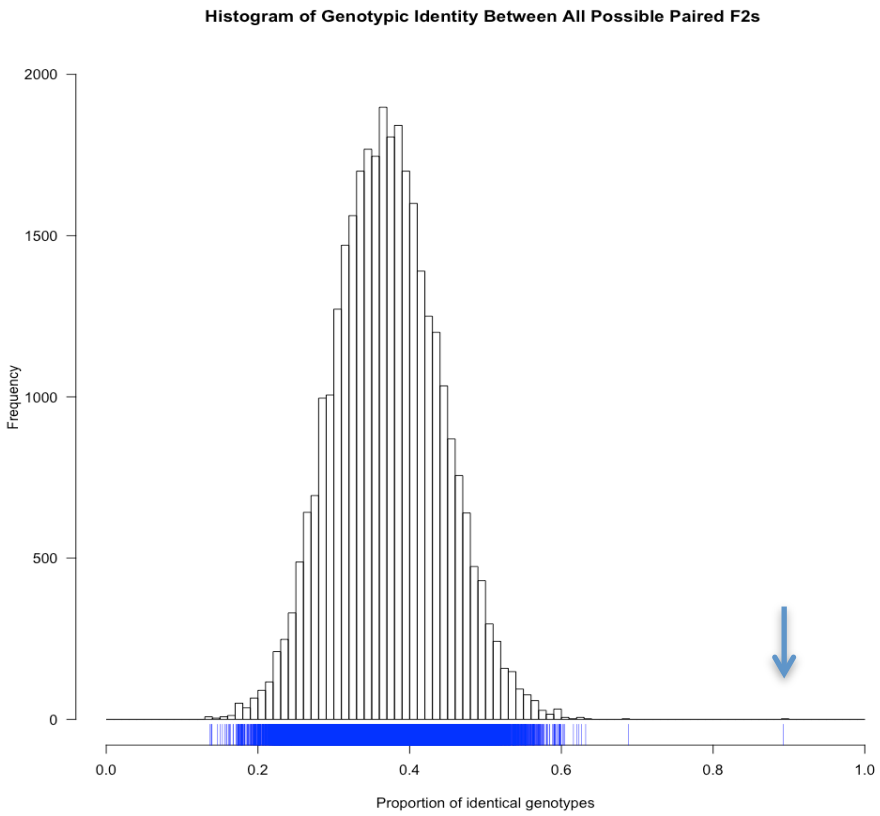

Figure 6 Degree of genotypic identity among all possible paired F2s in the RvT1 x Lche4 of 181 F2s genotyping for 1002 SNP markers

Figure 6 showed the distribution of all possible pairs of 181 F2s related to their paired marker genotypic identities. The R function comparegeno() was used to check identical or near identical pairs, which shared the same marker genotypes. This might suggest duplication sampling or DNA sample mixes between two F2s (Hwang, 2010). The majority of the pairs of individuals shared a range of 0.2~0.6, with a mean slightly less than 0.4. The reasons why a pair of F2 had a mean around 0.4 could be explained as follows. If one locus was considered for one pair of F2s, both individuals could be AA,

AB, BB, with a probability 0.25, 0.5, 0.25, respectively. The possible identical genotypes for two individuals were of 3 types: AA/AA, AB/AB, BB/BB. Therefore, the expected mean probability could be calculated as 0.25*0.25+0.5*0.5+0.25*0.25=0.375, similar to the measured value. Interestingly, there was one pair with genotypic identities around 0.9. The two individuals were further examined by using the clonestat () function. The individuals L4 and L5 shared 89.28% markers identity out of 858 markers. Whether they were random outliers or their DNA samples were mixed during the experiment procedures was not known. As a result, in the analysis, L4 and L5 were identified as suspicious individuals, whether they had detrimental effects was checked when constructing the linkage map.

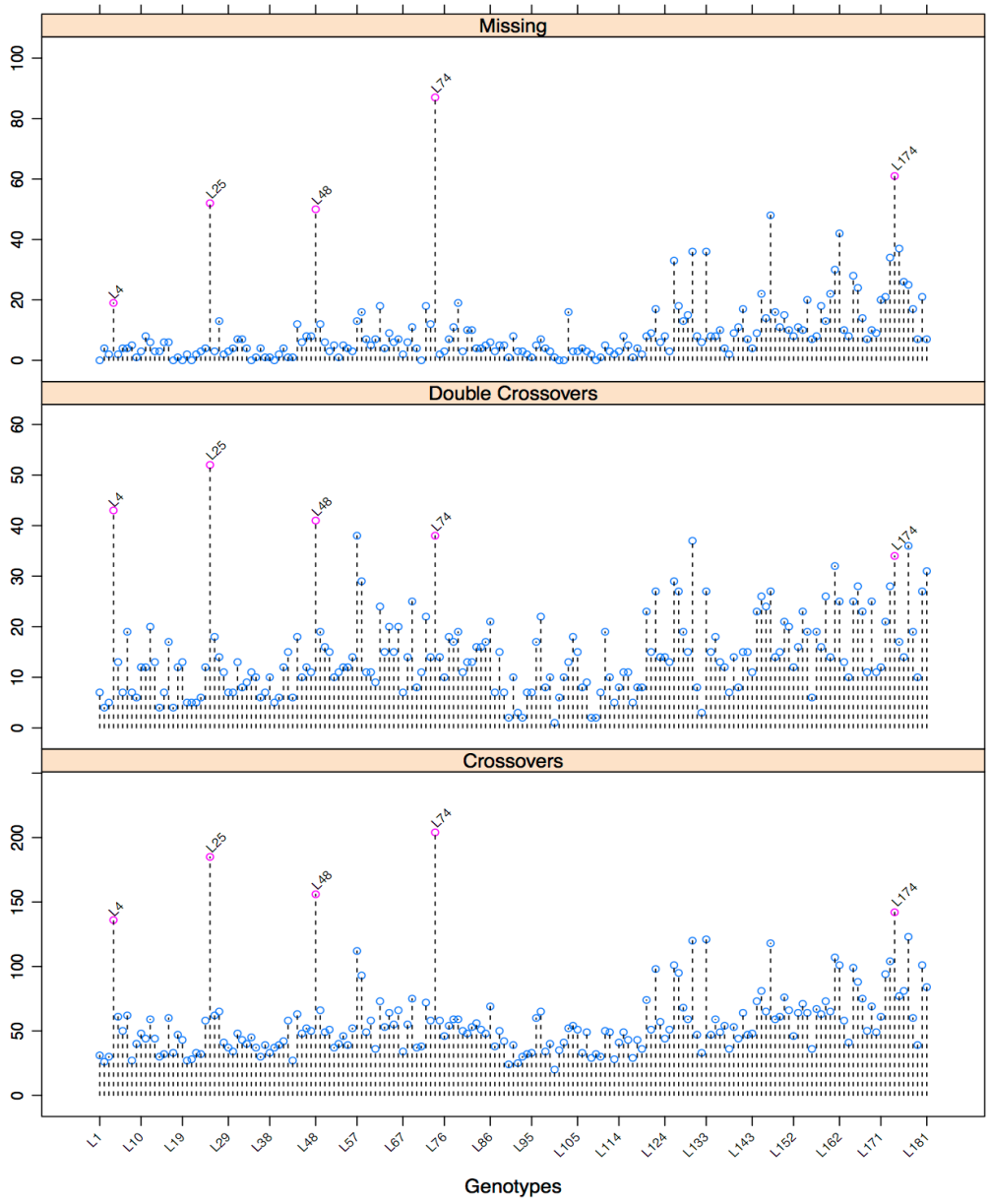

Figure 7 The missing values, double crossovers and crossovers for individual genotype

Figure 7 showed the missing values, the number of double crossovers and crossover for each of 181 individuals. No individual has a missing rate that exceeded

15%. The highest missing rate was 150/1002=14.9%. The majority of individuals had 10~40 double-crossover events and 50~150 crossover events. The range of double crossover rates per individual was from 0.01(10/1002) to 0.06(60/1002). The range of the crossover rate per individual was from 0.05 (50/1002) to 0.25(250/1002). Several individuals (red spots) with high missing rates tended to have more double crossovers and crossovers. As their missing rates didn't exceed 15%, we chose to keep those data but monitored them closely in the following analysis.

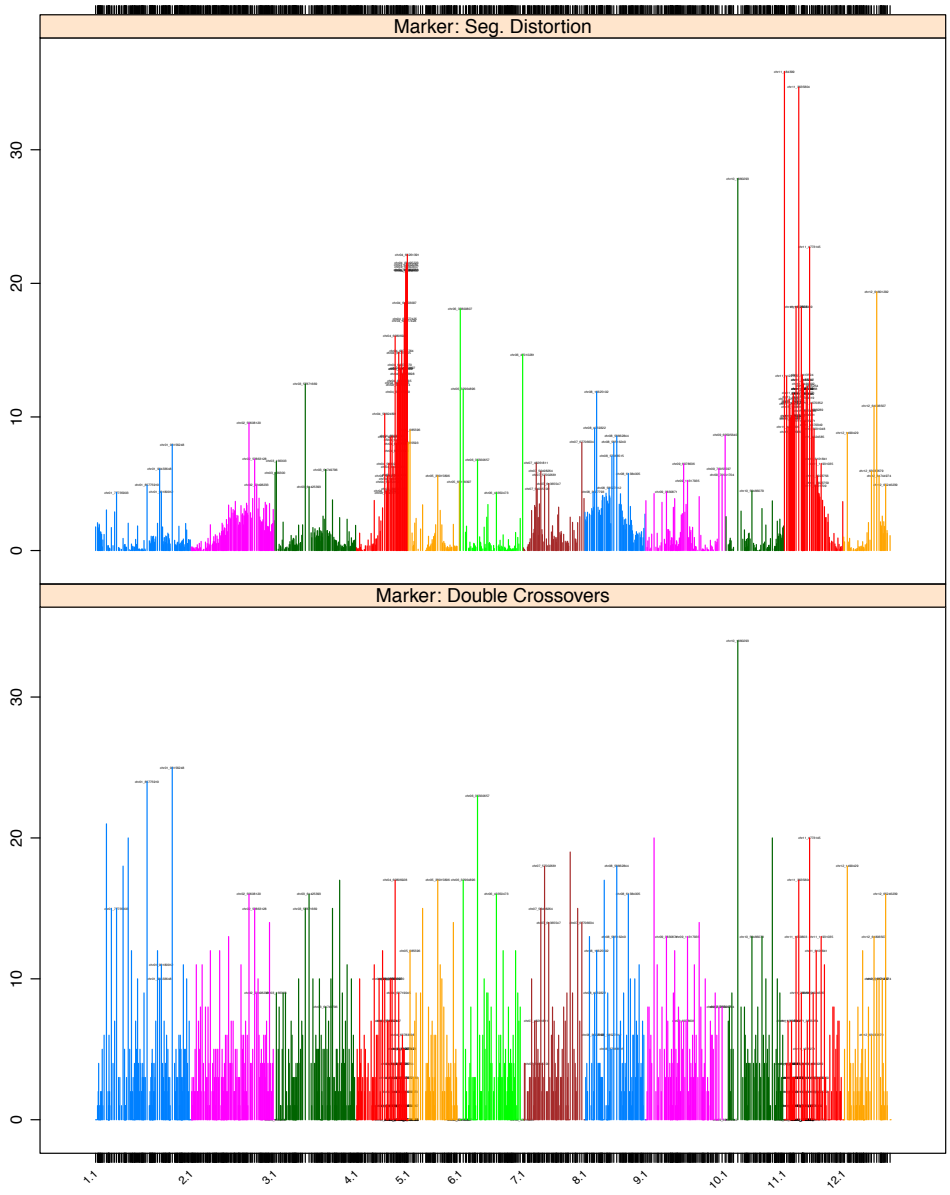

Figure 8 Marker profiles of the -log10 p-value for the test of segregation distortion and the number of double crossovers in the F2 population

The segregation distortion is a common phenomenon. It was occurs when allelic frequencies depart from the expected frequencies predicted by Mendelian law (Taylor & Butler, 2015). This distortion happens due to the inadvertent selection in the breeding process, or in the genomic regions from underlying biological and genetic mechanisms (Taylor & Butler, 2015). Figure 8 implied there was a spike of marker segregation distortions on chromosome 4 and chromosome 11. After examining the allelic frequencies of the segregation-distorted markers on chromosome 4 and chromosome 11, the B alleles from the male parent Lche4 showed a reduced frequency, which indicated real distortion. In a mapping study of RILs between *S. lycopersicum* and *S. cheesmaniae*, the strong segregation distortion on chromosome 4 and chromosome 11 had been observed in the vicinity of the gametophytic factor, which caused selective abortion of gametes containing Lche4 alleles (Paran et al., 1995).

However, it should not cause concern since multiple estimates of genetic map distances are little affected by real segregation. The real concern would be a single distorted marker surrounded with other markers that normally segregated, which is more possible to be a genotyping error (Broman & Sen, 2009). So we decided to keep the markers, which showed segregation distortion on chromosome 4 and 10. For the markers on the other chromosomes, we have removed the very skewed markers (-log10 P >10), and kept a close eye on the skewed markers ($-\log_{10}(P) \geq -\log_{10}(0.05/1002)=4.30$), which would be removed if they have deleterious effects on the linkage map. According to (Hyma et al., 2015), markers that resulted in double crossovers for more than 20% of progeny were recommended to be discarded. None of the markers had a double

crossover rate of 20% in the current results. "chr11_2567551" ,"chr11_2987399" and "chr11_2987402" were found to be duplicated and two of them would be deleted.

The linkage map

The first round selected markers were imported into JoinMap 4.1 to construct a linkage map. The suspicious markers were scrutinized one by one. If they increase the map distance, we removed them. Also for the markers that were too close together, like just 3 base pairs away in the physical position, the one, which expanded the map distance, was removed, because they were probably caused by sequencing duplicity.

Finally, 742 SNP markers were successfully mapped onto 12 chromosomes (Figure 9; Table 5) that spanned 1319.47 cM. The average distance between two adjacent markers was 1.78 cM, while the maximum gap between markers was 19.68 cM. Given the tomato genome size of about 950 Mb, the current genetic map was about 0.72 Mb/cM. Compared to the former linkage map constructed in a F2 population from *S. lycopersicum* and *S. cheesmaniae* with 71 RFLP markers, which spanned 1023 cM, this GBS map represented around a ten-fold improvement in resolution from 17.3 cM to1.78 cM between adjacent markers (Paterson et al., 1991). A more recent and comparable linkage map was constructed between *S. lycopersicum* and *S. galapagense* (Firdaus et al., 2013), using 589 SNP markers in a population of 182 F2s, which spanned 1259 cM with a resolution of 2.13 cM per marker. Our GBS map was a little longer than this map, but more precise in resolution because more markers had been mapped. Generally, our map was a medium-density map compared to other tomato linkage maps (*S. lycopersicum* × *S.*

pennelli), but might be the highest-density map for the interspecific population *S. lycopersicum* × *S. cheesmaniae.*

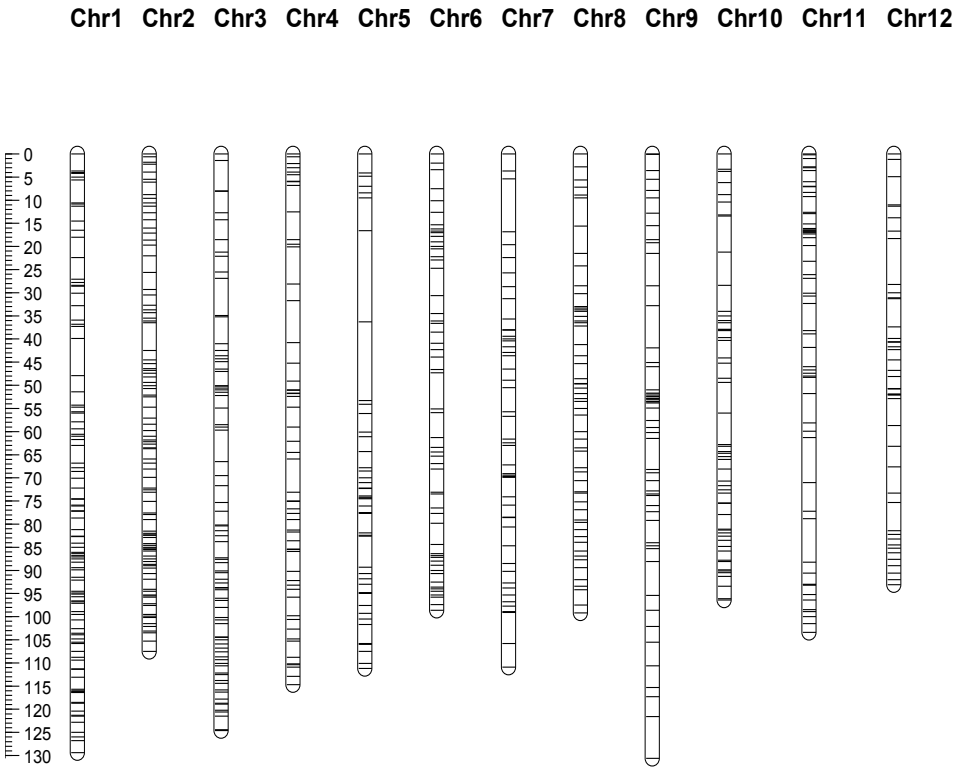

Figure 9 Linkage map of the F2 mapping population

Table 5 Distribution of SNP markers on the linkage map

Chr	No. of markers	Length (cM)	Average marker interval (cM)	Maximum Interval (cM)
Chr1	99	129.41	1.31	7.94
Chr2	95	107.47	1.13	5.91
Chr3	80	124.57	1.56	8.04
Chr4	56	114.68	2.05	9.11
Chr5	45	111.24	2.47	19.68
Chr6	56	98.56	1.76	7.82
Chr7	48	110.86	2.31	11.41
Chr8	56	99.22	1.77	6.17
Chr9	54	130.56	2.42	9.09
Chr10	54	96.39	1.79	7.73
Chr11	56	103.37	1.85	9.69
Chr12	43	93.14	2.17	9.90
Average	62	109.96	1.78	-
Total	**742**	**1319.47**		-

Heat map is a very good visual diagnostic tool to check the quality of the linkage map (Zhang et al., 2016). The heat map combines marker-to-marker linkage recombination fraction with the LOD scores together, which indicates the strength of linkage. Assuming that the estimate of recombination frequency between adjacent markers is r, the LOD score is to test for no linkage when r=0.5. Higher LOD scores meant the null hypothesis that no linkage between the two markers was rejected and the

possibility of linkage was strong (Taylor & Butler, 2015). Red spots indicated adjacent markers were linked (low r or high LOD), while blue spots indicated they are not linked (high r or low LOD) (Broman & Sen, 2009). A plot with a diagonal red line implied a good assignment with consistent marker order and a good quality linkage map (Lendenmann, Croll, Stewart, & McDonald, 2014).

Figure 10 showed consistent red spots across the diagonal line within chromosomes, indicating strong linkage between nearby markers and our linkage map had a good quality.

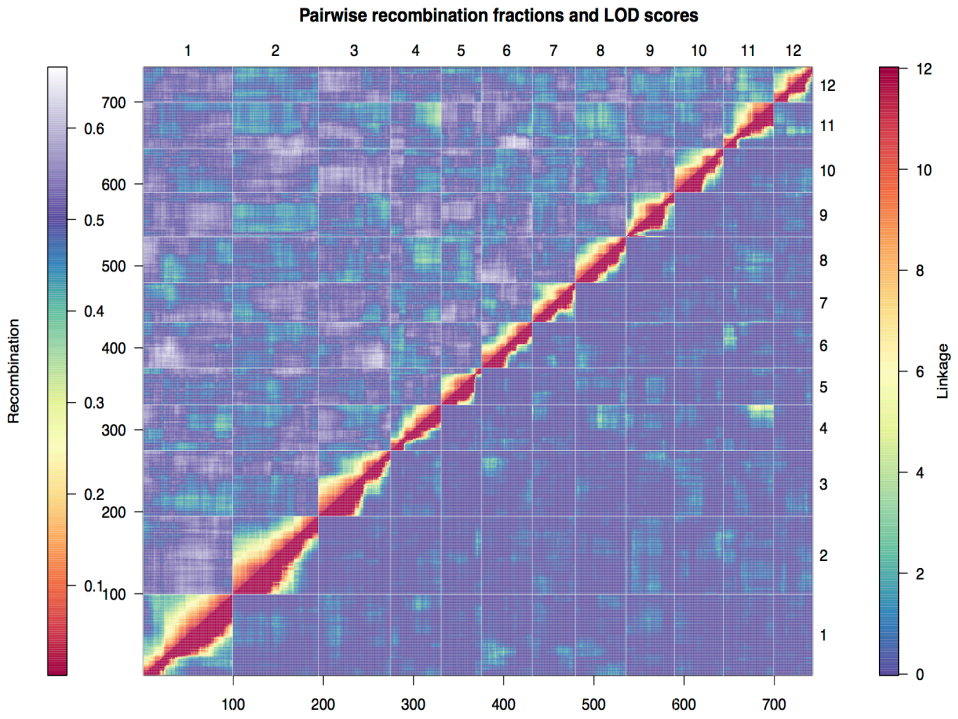

Figure 10 Heat map of constructed linkage map, plotted by markers assigned in chromosomes from 1 to 12

CHAPTER IV

QUANTITATIVE TRAIT LOCI MAPPING IN THE RVT1 X LCHE4 POPULATION

QTL analysis for complex traits is not an easy task, and depends heavily on complex statistical methods. Nowadays, with the evolution of statistical methods, more precise QTL locations can be detected. Single marker tests such as t-test, ANOVA test, and liner regression test are methods to test the means of phenotype differences corresponding to their genotype. When the sample data pass with a significant value, it proves that the marker is associated with the trait. The main drawback of the single marker test is that it can't give a closed form estimate for the QTL location (Doerge, 2002). An interval mapping method was proposed by Lander and Botstein (Lander & Botstein, 1989). This method used a statistical algorithm to scan the interval between the two adjacent markers on the linkage map, and followed by conducting a LOD test. If the LOD peak passed a threshold, there was a great possibility that a QTL existed (Gan, 2014). The disadvantage of interval mapping was the ignorance about the effect of other QTLs near the testing QTL. The estimated of QTL effects in the model was biased if there were many linked QTLs (Wang, Basten, & Zeng, 2011). Zeng and Jasen proposed an improved method called "Composite Interval Mapping" independently, which combined interval mapping and multiple regression methods together. The interval mapping was used to fit a linear model for every marker after a genome scan and

multiple regression was used to fit cofactors to control the linked and unlinked QTL effects and thus to reduce the model residual (Silva et al., 2014). However, the above methods are all single QTL analyses, which only can estimate single QTL effects, with ignorance of epistatic effects among QTLs. An upgraded mapping strategy called MCIM (mixed-model based composite interval mapping) was devised to evaluate putative locations of multiple QTLs and their interactions (Yang, Zhu, & Williams, 2007). A Bayesian method implemented with Gibbs sampling, estimated the genetic parameters in the mixed-full model that does a 2 dimensional genome searching for epistatic effects among QTLs (Yang et al., 2007). This method used a different F-statistic based on Henderson method III for a hypothesis test that was less computationally intensive than the likelihood ratio test. Another mapping strategy that needs to be mentioned is the nonparametric interval mapping, which specifically aims to map the phenotypes that don't follow a normal distribution. Nonparametric interval mapping was the extension of the Kruskal-Wallis test, in which there was an arbitrary number of genotype groups (Broman & Sen, 2009).

The objective of this chapter is to combine the phenotypic data with a well-constructed linkage map and use three QTL mapping methods, including composite interval mapping, nonparametric interval mapping and mixed-model based composite interval mapping, to search for QTLs associated with corresponding phenotypes.

Methods

The composite interval mapping was performed using the software WinQTL Cartographer v2.5 to detect QTLs for the 14 traits. Stepwise forward and backward

regression were analyzed using the standard model recommended by Wang et al. (2007), in which a window size of 10 cM, a walk speed of 1 cM and five control markers were adopted. The window size meant the neighboring regions on the left and right sides enclosing a putative QTL. Any cofactor marker located within the window size was excluded as it was considered to be too tightly linked to the putative QTL (Tijsterman et al., 2014). The threshold of LOD score for QTL detection was determined with 1000 permutations ($P < 0.05$), which could control the amount of false positives, known as Type I error (Churchill & Doerge, 1994).

The non-parametric interval mapping detecting QTLs for non-normal phenotypes (root length, root surface area, root volume, root dry weight, root fresh weight, root to shoot ratio, specific root length, root tissue density) was conducted with R/qtl. The threshold for QTL detection was also determined by a 1000 permutation test ($P<0.05$).

The epistatic effects among QTLs were examined with the software QTL Network 2.0. A one-dimensional genome scan was executed to identify the major QTLs and then a two-dimensional genome scan was used to test all possible combinations of two loci if they had an epistatic effect (Hwang et al., 2013). The F-test threshold was calculated by 1000 times permutation. The genome scan was done with a 10 cM testing window size, 10 cM filtration window size, at a walk speed of 1 cM.

Results and discussion

Table 6 QTLs detected for all the traits in the F2 population by WinQTLCart 2.5

Trait	Number	Nearest Marker	QTL Position	LOD Score	Additive Effect	Dominant Effect	R^2
SL	1	chr01_4205504	22.4	5.907	1.9505	1.095	5.09%
	2	chr02_43580135	61	4.3767	1.971	0.6592	4.98%
	3	chr06_25980660	17	5.9211	-2.2132	-0.3592	8.49%
	4	chr06_44717016	73.5	4.8735	1.5823	1.4055	1.94%
	5	chr11_3855254	25.2	5.5006	-1.635	1.6393	12.16%
	6	chr11_4670349	30.7	6.2183	-1.6125	1.6925	13.15%
SFW	1	chr04_1125394	6.8	4.4077	1.5434	0.3615	7.57%
RFW	1	chr01_3279061	17.5	5.3391	0.729	-0.5196	18.36%
	2	chr01_4205504	26.4	4.4608	0.7222	-0.0842	11.11%
RDW	1	chr01_3172890	15.5	6.071	0.0476	-0.0302	18.99%
	2	chr01_4205504	26.4	5.3881	0.0481	-0.0067	13.18%
	4	chr04_1002319	5.9	2.7868	0.0404	0.0175	4.95%
RAD	1	chr04_55053452	64.2	4.239	0.0219	-0.0093	12.06%
RS	1	chr03_67259598	106.8	4.1775	-26.2539	61.7122	9.72%
SPAD	1	chr01_3172890	14.5	8.4577	2.0361	1.3572	5.21%
	2	chr01_4205504	23.4	15.1831	2.659	1.9838	8.75%
	3	chr01_94270877	111.4	4.1874	1.6971	0.2022	5.39%
	4	chr01_96174801	121.3	4.7329	1.8131	0.5577	4.29%
	5	chr02_46803547	73.2	6.7082	2.1241	-0.1733	10.57%
	6	chr02_48414726	80	4.9483	1.7982	-0.1967	8.46%
	7	chr08_61195416	73	5.2792	1.8588	0.5154	5.65%

*Additive effect of the difference in the studied trait at a marker locus homozygous for RvT1 vs. homozygous for Lche4. A positive value indicated the favorable allele for increasing the trait value was inherited from RvT1. A negative value indicated the favorable allele for increasing the trait value was inherited from Lche4.

Generally, the identified QTL may be described as ‘major’ or ‘minor’. This definition is based on the proportion of the phenotypic variation explained by a QTL. Major QTL will account for a relatively large proportion (e.g. >10%), and minor QTL will usually account for <10% (Boopathi, 2013). In total, 20 significant QTLs were identified for 7 traits by composite interval mapping (Table 6). Six QTLs were identified on the chromosome 1, 2, 6, 11 for shoot length. Both parents contributed favorably to

shoot length, with RvT1 contributing three QTLs and Lche4 contributing three QTLs. Two major QTLs were on the chromosome 11, which explained 12.16% and 13.15% of phenotypic variance in shoot length, respectively, increasing the additive genetic effect of each loci by 1.635 cm and 1.613 cm (Mason, 2010). Two major QTLs on the chromosome 1 were identified for root fresh weight, explaining 18.36% and 11.11% variance, respectively. Similarly, two major QTLs on the chromosome 1 were associated with root dry weight, which explained 18.99% and 13.18% of the phenotypic variance, respectively. These QTLs for root fresh weight and root dry weight were due to favorable alleles from RvT1 (Warrington et al., 2015), which suggested that RvT1 was the major genetic donor for both traits. No significant QTLs were detected for specific root length. For the SPAD, 7 QTLs were identified on chromosome 1, 2, 8, with all the favorable alleles from parent RvT1. One major QTL (flanking by chr02_46803547) explained 10.57% of the phenotypic variance for SPAD.

Table 7 QTLs detected for non-normal traits in the F2 population by R/qtl

Trait	Nearest Marker	QTL position	1.5 LOD Support Interval(cM)	Peak LOD	Additive Effect	Dominant effect	R2
RL	chr04_3261417	31.7	28.10-40.77	4.46	296.36	352.76	7.70%
RSA	chr04_3261417	31.7	28.10-40.77	4.71	49.7	52.75	9.85%
RV	chr04_3261417	31.7	28.10-40.77	5.14	0.58	0.57	11.55%
RDW	chr01_5585767	26	0-47.89	4.08	0.04	-0.004	8.17%
	chr04_3261417	31.7	4.55-40.77	5.01	0.04	0.028	8.27%
RFW	chr01_5585767	26	0-32.80	3.87	0.62	-0.06	7.16%
	chr04_3261417	31.7	28.10-40.77	5.24	0.55	0.51	7.84%
SRL	chr01_89056634	91	35.91-118.46	3.85	-10.98	-6.39	8.48%
R:S	chr01_5585767	25	5.58-39.95	5.89	0.07	0.02	15.24%

*Additive effect of the difference in the studied trait at a marker locus homozygous for RvT1 vs. homozygous for Lche4. A positive value indicated the favorable allele for increasing the trait value was inherited from RvT1. A negative value indicated the favorable allele for increasing the trait value was inherited from Lche4.

For the phenotypes that didn't follow a normal distribution, the nonparametric interval mapping method was used to successfully detect corresponding QTLs, except for root tissue density. There was one common QTL (flanking by chr04_3261417) on the chromosome 4 associated with root length, root surface, root volume, root dry weight and root fresh weight. This common QTL was detected probably because these five traits were highly significantly correlated with each other. A QTL (flanked by chr01_89056634) on the chromosome 1 was detected for specific root length, which explained 8.48% of the phenotypic variance. The additive effect of the QTL was -10.98, and the dominant effect was -6.39. This indicated that the Lche4 alleles of the QTL increased the specific root length, which was reduced when the Lche4 allele was replaced by the allele from the cultivated tomato RvT1 (Chen et al., 2014).

Table 8 QTLs detected for all the traits in the F2 population by QTL Network

Trait	QTL interval	QTL Position	Position Range	Additive Effect	Dominant Effect	R2
SL	chr11_1970271-chr11_2164281	12.9	10.2-14.9	-1.2062	2.4153	12.70%
SPAD	chr01_3572627-chr01_4205504	21	17.4-32.1	2.8926	1.3629	20.01%
	chr02_46803547-chr02_47484815	73.2	72.6-74.2	2.0774	-0.0777	29.97%
RFW	chr04_3261417-chr04_4545307	31.7	30.1-35.7	0.617	0.6266	10.65%
RDW	chr01_3279061-chr01_3572627	17.4	5.6-21.0	0.038	-0.0264	10.77%
	chr04_3261417-chr04_4545307	31.7	29.1-36.7	0.037	0.0302	9.65%
R:S	chr01_2308792-chr01_2751392	19	16.4-22.4	0.0723	-0.0425	11.69%
RS	chr04_3261417-chr04_4545307	31.7	30.1-35.7	49.6525	52.8427	9.85%
RV	chr04_3261417-chr04_4545307	31.7	29.1-36.7	0.5802	0.5681	11.56%

*Additive effect of the difference in the studied trait at a marker locus homozygous for RvT1 vs. homozygous for Lche4. A positive value indicated the favorable allele for increasing the trait value was inherited from RvT1. A negative value indicated the favorable allele for increasing the trait value was inherited from Lche4.

The QTL network was aimed at examining the epistatic effect between markers, but we couldn't find any epistatic effect among QTLs. Two QTLs with large effect were detected for SPAD. One QTL (flanked by chr01_3572627-chr01_4205504) accounting for 20.01% phenotypic variance, and the other one QTL (chr02_46803547-chr02_47484815) accounting for 29.97%. One major QTL (flanked by chr01_2308792-chr01_2751392) was identified for the root-to-shoot ratio, explaining 11.69% of the phenotypic variation. The QTLs associated with root fresh weight, root dry weight, root surface and root volume detected in nonparametric interval mapping were also found with QTL Network 2.0.

SUMMARY

In total, 27 QTLs have been identified for the different root and shoot traits on chromosomes 1,2,3,4,6,8, and 11. The QTLs for most root traits were located on the chromosome 1 and 4. The results of our study suggest that root morphological traits in our population are regulated by a suite of small-effect-loci and several major effect loci. One major QTL (flanked by chr04_3261417) associated with root length, root surface area, root volume, root fresh weight and root dry weight was identified. This highlighted QTL probably could be a useful marker to screen the traits simultaneously. The parent Lche4 has been characterized as a major genetic donor of high specific root length, and could be a useful germplasm to modify the root system of cultivated tomato. All the identified QTLs need to be verified in different environments in the future to verify stable QTLs, because root traits are usually affected by environmental factors, like different soil types, fertility, temperature and water stress.

The SNP markers discovered in this study should be useful for root system improvement in tomato, and the linkage map constructed for *S. lycopersicum and S. cheesmaniae* could provide more genetic information for other breeders who want to utilize the wild species *S. cheesmaniae* into their breeding programs for drought tolerance, whiteyfly resistance or a increase of fruit soluble solids.

REFERENCES

Ashrafi Nasrabadi, H. (2007). *QTL mapping of early blight resistance and fruit quality related traits in a Lycopersicon esculentum x L. pimpinellifolium RIL population of tomato* (Doctoral dissertation). Pennsylvania State University, College Park, PA.

Bai, Y., & Lindhout, P. (2007). Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Annals of botany*, *100*(5), 1085–1094.

Bauchet, G., & Causse, M. (2012). Genetic diversity in tomato (*Solanum lycopersicum*) and its wild relatives. In Muhammed, A. (Ed.), *Genetic diversity in plants* (pp. 133-162). Rijeka, Croatia: InTech

Bergougnoux, V. (2014). The history of tomato: From domestication to biopharming. *Biotechnology advances*, *32*(1), 170–189.

Boopathi, N. M. (2013). *Genetic mapping and marker assisted selection: basics, practice and benefits*. New Delhi, India: Springer India

Boriss, H., & Brunke, H. (2005). Commodity profile: tomatoes, fresh market. Available from: http://aic.ucdavis.edu/profiles/FreshTomatoes-2005.pdf, accessed on 09/17/2015.

Boswell, V. R. (1952). Plant breeding and the vegetable industry. *Economic botany*, *6*(4), 315-341.

Broman, K. W. (2001). Review of statistical methods for QTL mapping in experimental crosses. *Lab animal*, *30*(7), 44–52.

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl* (Vol. 46). New York, NY: Springer.

Chen, A. L., Liu, C. Y., Chen, C. H., Wang, J. F., Liao, Y. C., Chang, C. H., ... & Chen, K. Y. (2014). Reassessment of QTLs for late blight resistance in the tomato accession L3708 using a restriction site associated DNA (RAD) linkage map and highly aggressive isolates of *Phytophthora infestans*. *Plos one*, *9*(5), e96417.

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, *138*(3), 963–971.

Cortina, C., & Culiáñez-Macià, F. A. (2005). Tomato abiotic stress enhanced tolerance by trehalose biosynthesis. *Plant science*, *169*(1), 75-82.

Crespi, M. (Ed.). (2013). *Root genomics and soil interactions*. Danvers, MA: Wiley-Blackwell

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature reviews genetics*, *3*(1), 43–52.

Dorais, M., Papadopoulos, A. P., & Gosselin, A. (2001). Greenhouse tomato fruit quality. *Horticultural reviews*, *26*(2001), 239-319.

Eissenstat, D. M. (1991). On the relationship between specific root length and the rate of root proliferation: a field study using citrus rootstocks. *New phytologist*, *118*(1), 63–68.

Falconer, D.S., Mackay, T.F.C. (1996). *Introduction to quantitative genetics.* Harlow, UK: Longman Group Limited

Firdaus, S., van Heusden, A. W., Hidayati, N., Supena, E. D. J., Mumm, R., de Vos, R.

C. H., … Vosman, B. (2013). Identification and QTL mapping of whitefly resistance components in *Solanum galapagense*. *Theoretical and applied genetics*, *126*(6), 1487–1501.

Foolad, M. R. (2007a). Current status of breeding tomatoes for salt and drought tolerance. In *Advances in molecular breeding toward drought and salt tolerant crops* (pp. 669-700). Dordrecht, The Netherlands: Springer Netherland

Foolad, M. R. (2007b). Genome mapping and molecular breeding of tomato. *International journal of plant genomics*, *2007*, 1-52

Foolad, M. R., Ntahimpera, N., Christ, B. J., & Lin, G. Y. (2000). Comparison of field, greenhouse, and detached-leaflet evaluations of tomato germplasm for early blight resistance. *Plant disease*, *84*(9), 967-972.

Foolad, M. R., & Panthee, D. R. (2012). Marker-assisted selection in tomato breeding. *Critical reviews in plant sciences*, *31*(2), 93-123.

Fulton, T. M., Chunwongse, J., & Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant molecular biology reporter*, *13*(3), 207–209.

Girapunthong, N., VanSickle, J. J., & Renwick, A. (2003). Price asymmetry in the United States fresh tomato market. *Journal of food distribution research*, *34*(3), 51-59.

Gan, S. T. (2014). *The development and application of molecular markers for linkage mapping and quantitative trait loci analysis of important agronomic traits in oil palm (Elaeis guineensis Jacq.)* (Doctoral dissertation). University of Nottingham,

University Park, Nottingham, UK

Ezura, H., Ariizumi, T., Garcia-Mas, J., & Rose, J. (Eds.). (2016). *Functional genomics and biotechnology in Solanaceae and Cucurbitaceae crops* (Vol. 70). Berlin, Heidelberg: Springer-Verlag.

Hackett, C. A., & Broadfoot, L. B. (2003). Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity*, *90*(1), 33-38.

Hartz, T., Miyao, G., Mickler, J., Le Strange, M., Stoddard, S., Nunez, J., & Aegerter, B. (1997). Processing tomato production in California. Available from https://anrcatalog.ucanr.edu/pdf/7228.pdf, accessed on 8/31/2015.

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in plant science*, *5*, 484.

Heuvelink, E. (Ed.). (2005). *Tomatoes* (Vol. 13). Cambridge, MA: CABI Publishing

Hochmuth, G. J. (2001). Fertilizer management for greenhouse vegetables. Available from http://edis.ifas.ufl.edu/cv265, accessed on 9/1/2015.

Hwang, S. (2010). *Linkage & selective mapping and QTL analysis in a Harosoy x Clark population of 300 RILs* (Doctoral dissertation). University of Nebraska-Lincoln, Lincoln, NE.

Hwang, S., King, C. A., Davies, M. K., Ray, J. D., Cregan, P. B., & Purcell, L. C. (2013). QTL analysis of shoot ureide and nitrogen concentrations in soybean (Glycine max (L.) Merr.). *Crop science*, *53*(6) 2421–2433.

Hyma, K. E., Barba, P., Wang, M., Londo, J. P., Acharya, C. B., Mitchell, S. E., ... & Cadle-Davidson, L. (2015). Heterozygous mapping strategy (HetMappS) for high resolution Genotyping-By-Sequencing markers: a case study in grapevine. *Plos one*, *10*(8), e0134880.

Jenkins, J. A. (1948). The origin of the cultivated tomato. *Economic botany*, *2*(4), 379-392.

Ji, Y., Scott, J. W., Hanson, P., Graham, E., & Maxwell, D. P. (2007). Sources of resistance, inheritance, and location of genetic loci conferring resistance to members of the tomato-infecting begomoviruses. In Czosnek, H. (Ed.), *Tomato yellow leaf curl virus disease* (pp. 343-362). Dordrecht, The Netherlands: Springer Netherland

Kalloo, G. (Ed.). (1991). *Genetic improvement of tomato* (1st ed.). Berlin, Heidelberg: Springer-Verlag.

Krieger, U., Lippman, Z. B., & Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature genetics*, *42*(5), 459–463.

Kuijken, R. C. P., Van Eeuwijk, F. A., Marcelis, L. F. M., & Bouwmeester, H. J. (2015). Root phenotyping: From component trait in the lab to breeding. *Journal of experimental botany*, *66*(18), 5389–5401.

Lander, E. S., & Botstein, S. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, *121*(1), 185-199

Lendenmann, M. H., Croll, D., Stewart, E. L., & McDonald, B. A. (2014). Quantitative

trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici. G3: Genes| Genomes| Genetics*, *4*(12), 2519-2533.

Liedl, B. E., Labate, J. A., Stommel, J. R., Slade, A., & Kole, C. (Eds.). (2013). *Genetics, genomics, and breeding of tomato*. Boca Raton, FL: CRC Press.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., … Huang, S. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nature genetics*, *46*(11), 1220–1226.

Mason, R. E. (2010). *Phenotypic and molecular genetic analysis of reproductive stage heat tolerance in wheat (Triticum aestivum)* (Doctoral dissertation). Texas A&M University, College Station, TX.

Mehra, M., Gangwar, I., & Shankar, R. (2015). A deluge of complex repeats: The *Solanum* genome. *Plos one*, *10*(8), e0133962.

Miller, J. C., & Tanksley, S. D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theoretical and applied genetics, 80*(4), 437–448.

Napier, A. B. (2006). *Identification and confirmation of molecular markers and orange flesh color associated with major QTL for high beta-carotene content in muskmelon* (Master's thesis). Texas A&M University, College Station, TX

Nuez, F., Prohens, J., & Blanca, J. M. (2004). Relationships, origin, and diversity of Galapagos tomatoes: implications for the conservation of natural populations. *American journal of botany*, *91*(1), 86-99.

Paez-Garcia, A., Motes, C., Scheible, W.-R., Chen, R., Blancaflor, E., & Monteros, M. (2015). Root traits and phenotyping strategies for plant improvement. *Plants*, *4*(2), 334–355.

Panguluri, S. K., & Kumar, A. A. (Eds.). (2013). *Phenotyping for plant breeding: applications of phenotyping methods for crop improvement*. New York, NY: Springer.

Paran, I., Goldman, I., Tanksley, S. D., & Zamir, D. (1995). Recombinant inbred lines for genetic mapping in tomato. *Theoretical and applied genetics, 90*(3–4), 542–548.

Paterson, A. H., Damon, S., Hewitt, J. D., Zamir, D., Rabinowitch, H. D., Loncoln, S. E., … Tanksley, S. D. (1991). Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. *Genetics*, *127*(1), 181–197.

Peralta, I. E., Knapp, S., & Spooner, D. M. (2005). New species of wild tomatoes (*Solanum* section *Lycopersicon*: Solanaceae) from Northern Peru. *Systematic botany*, *30*(2), 424-434.

Peralta, I. E., Knapp, S., & Spooner, D. M. (2006). Nomenclature for wild and cultivated tomatoes. *Tomato genetics cooperative report*, *56*, 6-12.

Robertson, L. D., & Labate, J. A. (2007). Genetic resources of tomato (*Lycopersicon esculentum* Mill.) and wild relatives. *Genetic improvement of Solanaceous crops-tomato, 2*, 25-75.

Ryser, P. (2006). The mysterious root length. *Plant and soil*, *286*(1), 1-6.

Shabala, S., & Munns, R. (2012). Salinity stress: physiological constraints and adaptive mechanisms. *Plant stress physiology*, *1*, 59-93.

Sim, S. C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Van Deynze, A., ... Francis, D. M. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *Plos one*, *7*(7), e40563.

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., … McCouch, S. (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theoretical and applied genetics*, *126*(11), 2699–2716.

Stevens, M. A., & Rick, C. M. (1986). Genetics and breeding. In Atherton, J. G. & Rudich, J. (Ed.), *The tomato crop* (pp. 35-109). Dordrecht, The Netherlands: Springer Netherland

Taylor, J., & Butler, D. (2015). ASMap: linkage map construction using the MSTmap algorithm. *R Package version 0.4-4*, http://CRAN.R-project.org/package=ASMap, accessed on 08/17/16.

Tijsterman, M., Okihara, K. L., Thijssen, K., Plasterk, R. H. a, Zhang, Y. Y., Sulston, J. E., … Bender, A. (2014). Composite interval mapping and multiple interval mapping: procedures and guidelines for using windows QTL cartographer. *Genetics*, *4*(1), 692–699.

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*(7400), 635-641.

Uga, Y., Sugimoto, K., Ogawa, S., Rane, J., Ishitani, M., Hara, N., … Yano, M. (2013).

Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nature genetics*, *45*(9), 1097–1102.

Villalta, I., Reina-Sánchez, A., Cuartero, J., Carbonell, E. A., & Asins, M. J. (2005). Comparative microsatellite linkage analysis and genetic structure of two populations of F6 lines derived from *Lycopersicon pimpinellifolium* and *L. cheesmanii. Theoretical and applied genetics*, *110*(5), 881–894.

Víquez-Zamora, M., Vosman, B., van de Geest, H., Bovy, A., Visser, R. G. F., Finkers, R., & van Heusden, A. W. (2013). Tomato breeding in the genomics era: insights from a SNP array. *BMC genomics*, *14*, 354.

Wang, S. C. J. B., Basten, C. J., & Zeng, Z. B. (2007). Windows QTL cartographer 2.5. Available from http://statgen.ncsu.edu/~shchwang/WQTLCart.htm, accessed on 08/15/16.

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., ... Boerma, H. R. (2015). QTL for seed protein and amino acids in the Benning× Danbaekkong soybean population. *Theoretical and applied genetics*, *128*(5), 839-850.

Wells, H. F., Bond, J., & Thornsbury, S. (2015). Vegetables and pulses outlook. Available from https://www.ers.usda.gov/media/1834600/vgs-355.pdf, accessed on 8/31/15.

Yang, J., Zhu, J., & Williams, R. W. (2007). Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics*, *23*(12), 1527–1536.

Zhang, D., Li, H., Wang, J., Zhang, H., Hu, Z., Chu, S., ... Yu, D. (2016). High-density

genetic mapping identifies new major loci for tolerance to low-phosphorus stress in soybean. *Frontiers in plant science*, *7*, 372.

APPENDIX

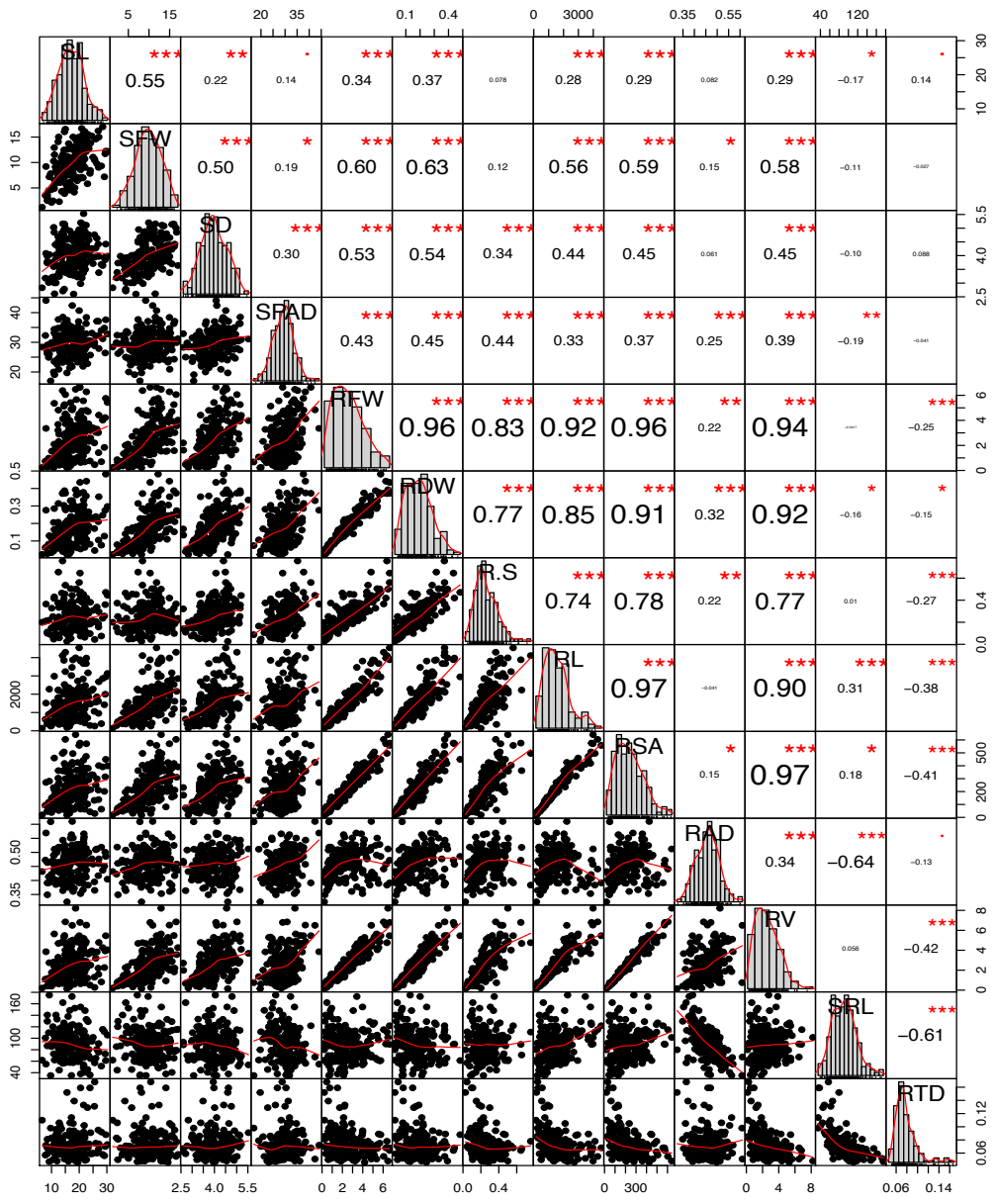

Figure 11 A combined scatter, histogram, and correlation coefficients plot

VITA

Name: Limeng Xie

Address: Horticultural Science Department, Texas A&M University

E-mail: momosan@tamu.edu, momosan@yahoo.com

Education: BS, Agronomy, China Agricultural University, Beijing, China, 2014